ESCAP MEETING NO. 24 - 12/20/00

AGENDA

Kathleen P Porter
12/20/2000 09:13 AM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC

cc: Vincent T Mule Jr/DSSD/HQ/BOC@BOC, Richard A Griffin/DSSD/HQ/BOC@BOC, Annetta Clark Smith/POP/HQ/BOC@BOC

Subject: ESCAP Agenda

The agenda for the December 20 ESCAP Meeting scheduled from 12:00-1:30 in Rm. 2412/3 is as follows:

Weight trimming results - Tom Mule

SBE - Rick Griffin and Annetta Clark-Smith

Voting Rights Act - Howard Hogan

ESCAP MEETING NO. 24 - 12/20/00

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

December 20, 2000

Weight Trimming Summary

What's weight trimming?

- The reduction of the sampling weights for clusters which could have extreme influence on the dual system estimates.

Why do weight trimming?

We trim weights

- To protect against undue influence from a small fraction of the sample, the outlier clusters and to reduce sampling variance.
- To reduce the effect of outlier clusters on post-stratum estimates.

What's the measure of cluster influence?

- The estimated **cluster** net error: the absolute difference between the estimated omissions and erroneous enumerations.
- $\text{Net Error} = |\text{omissions} - \text{erroneous enumerations}|$
- Estimates reflect:
  - Person match results
  - Sampling weights
  - Targeted Extended Search
  - Simple A.C.E. household noninterview adjustment
  - Simple imputation for unresolved status
- The 1990 PES used net error to measure cluster influence.

Which clusters are weight trimmed?

- If the estimated net error of a cluster exceeds a maximum pre-specified level, the sampling weights for that cluster are trimmed, or reduced.
- The maximum pre-specified levels are
  - American Indian Reservation 6,250
  - Balance of the U.S. 75,000
  - Puerto Rico 16,500

How were outlier clusters handled in 1990?

- For the 1,392 post-stratification, two small clusters were trimmed to a net error of approximately 150,000.
- For the 357 post-stratification, 104 clusters were rematched with the surrounding block search extended. After rematching, the largest net error was approximately 150,000. No weight trimming was done.

How are the weights reduced?

- A trimming factor is applied to the weights which reduces the cluster net error to the maximum pre-specified level.
- $$\text{Trimming Factor} = \frac{\text{Maximum Net Error}}{\text{Cluster Net Error}}$$

How many clusters were trimmed?

- 4 clusters were trimmed.
    - 1 cluster in the United States
    - 3 clusters in Puerto Rico
- No clusters on American Indian Reservation needed to be trimmed.

Table 1: Trimmed Clusters

Location	Cluster	Net Error Before Tramming	Net Error After Trimming	Weight Reduction Factor
Balance of the U.S.	1	77,975	75,000	3.8%
Puerto Rico	2	18,121	16,500	8.9%
Puerto Rico	3	16,712	16,500	1.3%
Puerto Rico	4	21,622	16,500	23.7%

Were the clusters that were trimmed outliers?

Yes.

Figures 1 - 4 show

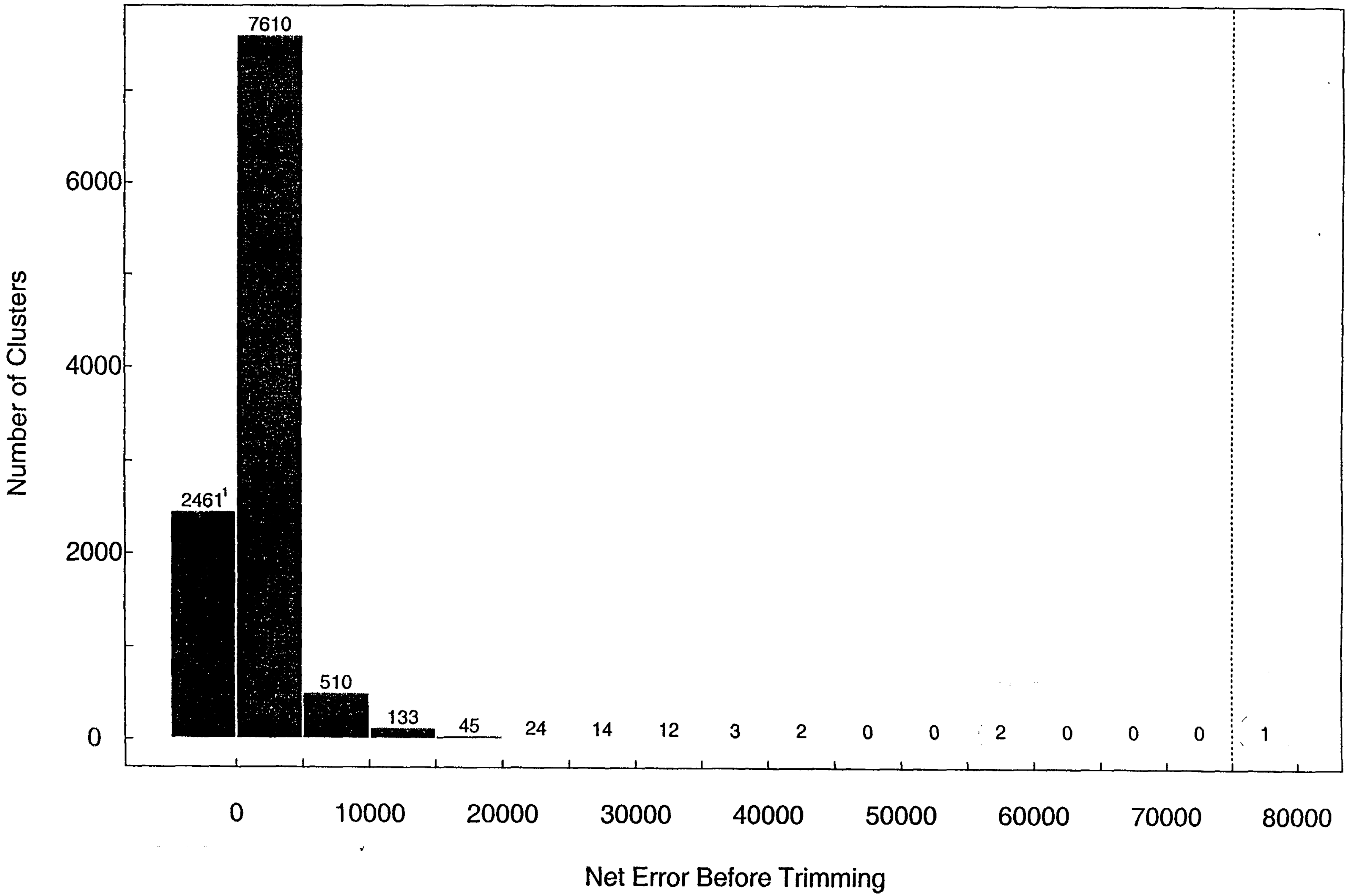
- The net error distributions are highly skewed.
- The clusters trimmed are at the extreme tails of the distributions.
- The shape of the net error distributions from weight trimming is very similar to the 1990 PES.

How did the trimming affect the estimates from weight trimming?

Weight trimming produced a very minimal change in estimates for the United States and Puerto Rico.

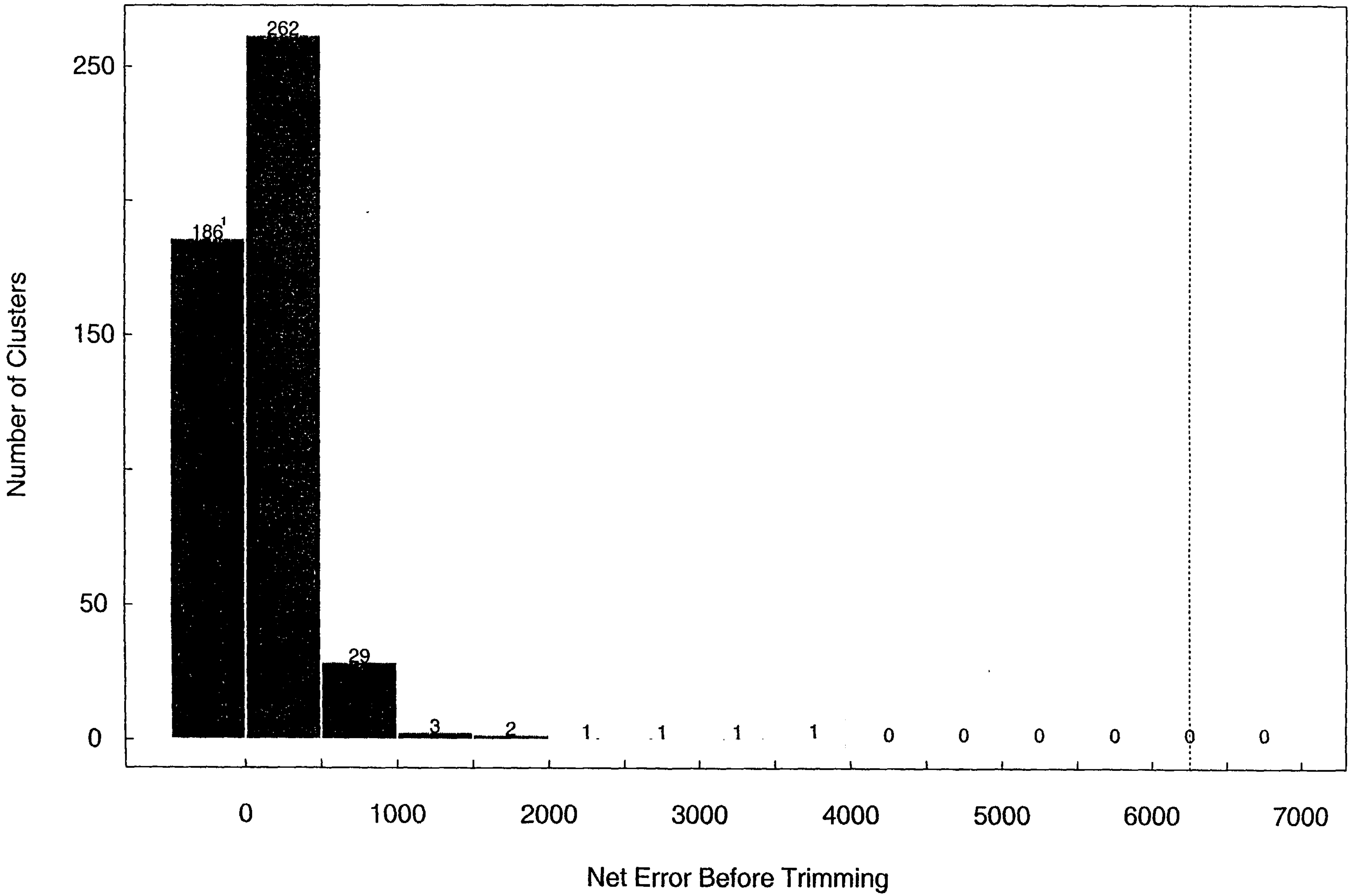
	Estimate	Before Trimming	After Trimming	Difference
United States	Weighted P-sample People	260,745,172	260,743,610	-1,562
	Weighted E-sample People	264,583,668	264,578,877	-4,791
Puerto Rico	Weighted P-sample People	3,460,491	3,452,677	-7,814
	Weighted E-sample People	3,534,125	3,533,700	-425

Figure 1. Distribution of 2000 A.C.E. Net Error for the Balance of the United States



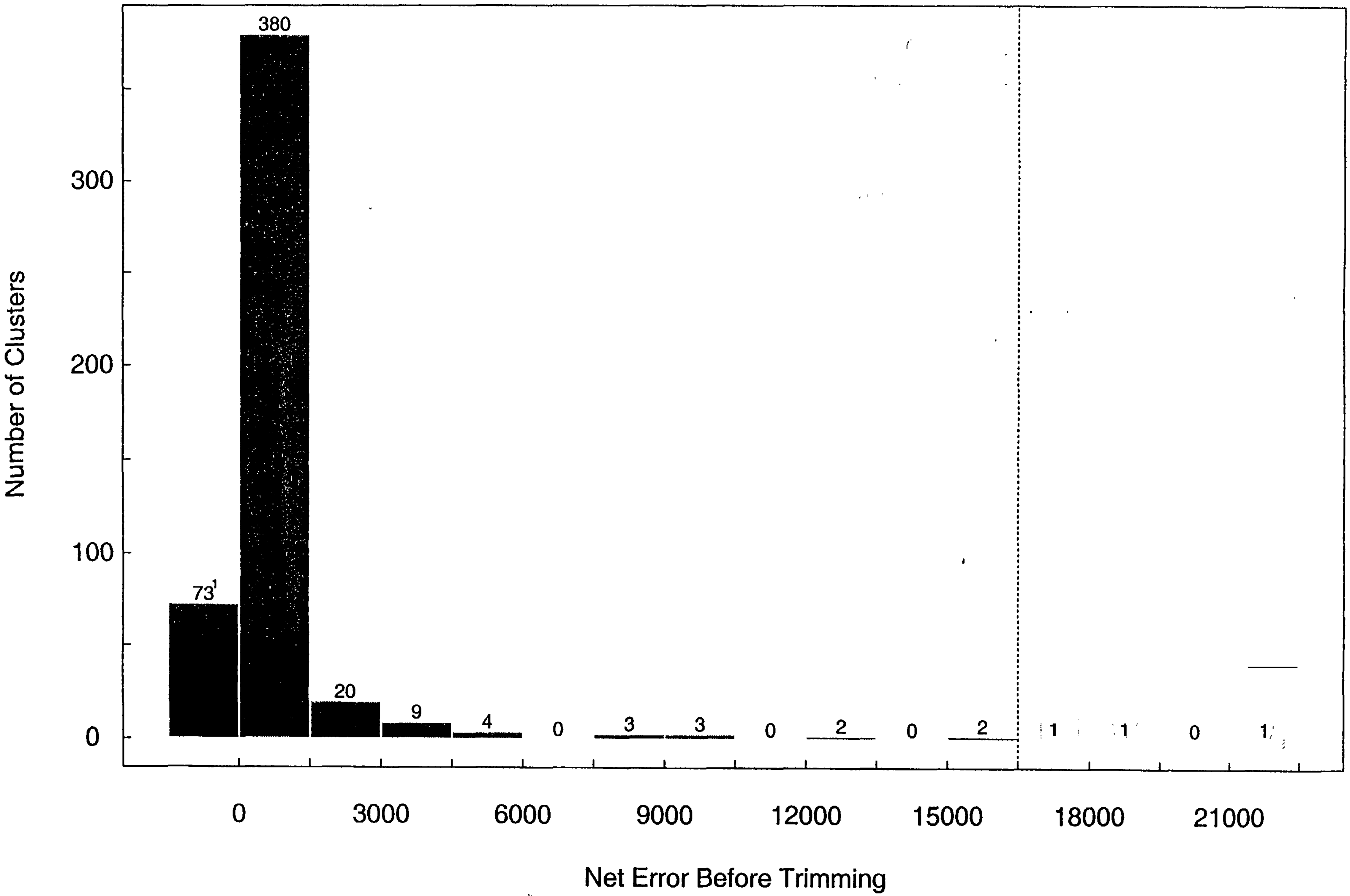
1 Clusters in this group have a net error exactly equal to zero

Figure 2. Distribution of 2000 A.C.E. Net Error for American Indian Reservations



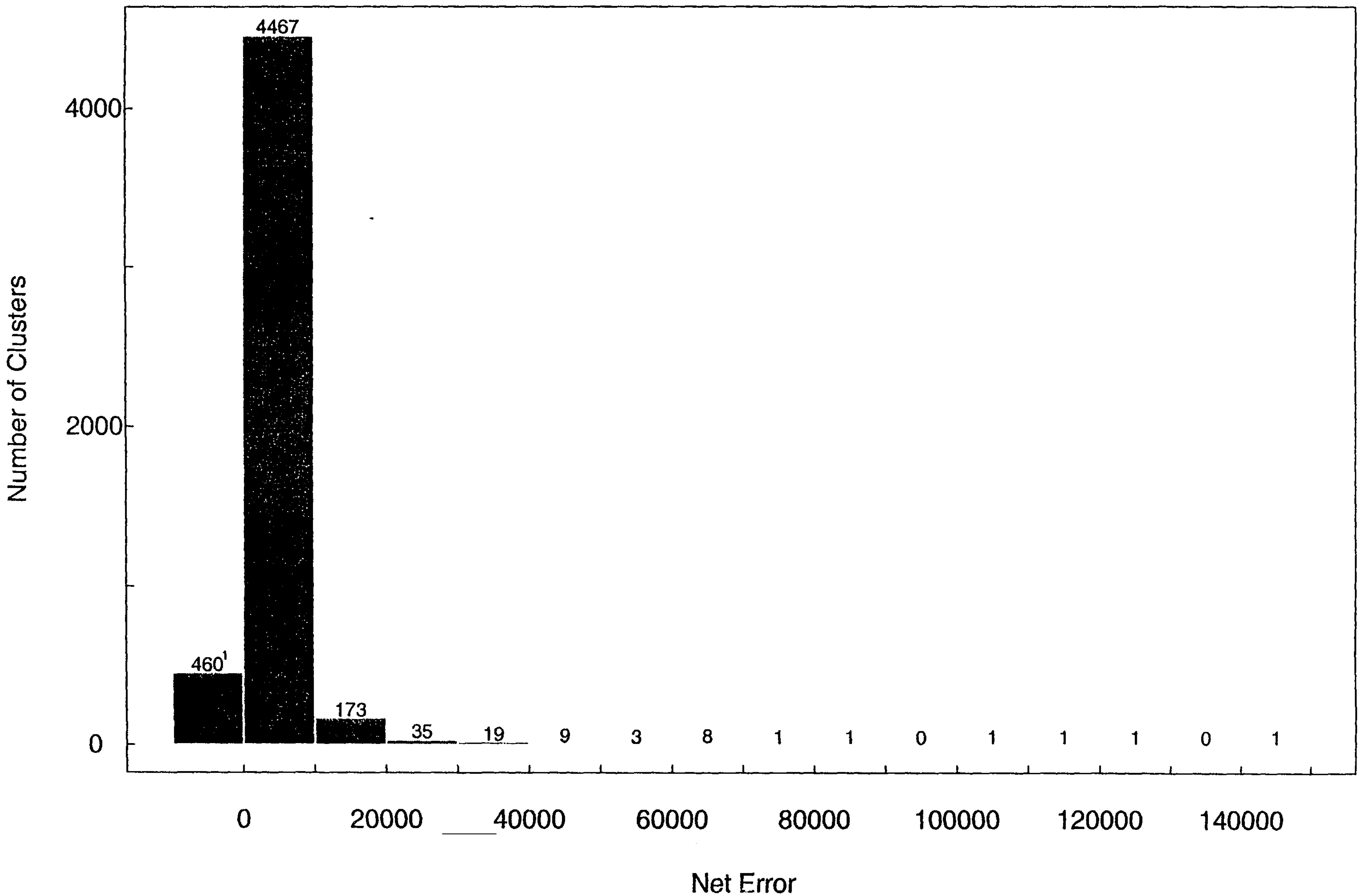
1 Clusters in this group have a net error exactly equal to zero

Figure 3. Distribution of 2000 A.C.E. Net Error for Puerto Rico



1 Clusters in this group have a net error exactly equal to zero

Figure 4. Distribution of 1990 PES Net Error



1 Clusters in this group have a net error exactly equal to zero

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

NSHAPC Results and Census 2000

Program Type	Number of Homeless Assistance Programs (National Survey of Homeless Assistance Providers and Clients - NSHAPC)	Number of Units (Census 2000)	Number of Program Contacts Expected on an Average Day in February 1996 (NSHAPC)	Number of People **Tabulated** (Census 2000)
Emergency Shelter Programs	5,700		240,000	
Transitional Housing Programs	4,400		160,000	
Voucher Distribution Programs	3,100		70,000	
Total Housing Programs	13,200	7,371	470,000	183,414
Soup Kitchen Programs	3,500	1,985	520,000	70,604
Mobile Food Van Programs	500	167	50,000	3,429
Total Food Programs	4,000	2,152	570,000	74,033

A homeless assistance program is a set of services offered to the same group of clients at a single location

NSHAPC offer an important overview of the service utilization in the United States. Since many people may use more than one type of service during an average day. the estimates of service levels made by NSHAPC programs necessarily contain an unknown and unknowable amount of duplication The results cannot be added up to determine the total number of clients who use services on an average day. For that reason they are referred to as "program contacts," not as "clients served."

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Comparison of Census 2000 and 1990
Service Locations

Program Type	Number of Units (Census 2000)	Number of Units (1990 Census)	PCT DIFF	Number of People Tabulated (Census 2000)	Number of People Tabulated (1990 Census)	PCT Diff
Emergency and Transitional Shelters	7,371	8,461	-12.88	183,414	190,406	-3.67
Outdoor Locations	4,701	6,669		23,080	49,734	
Soup Kitchen Programs	1,985			70,604		
Mobile Food Van Programs	167			3,429		
Total	14,224	15,130	-5.99	280,527	240,140	16.82

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

December 20, 2000 DRAFT

SERVICE BASED ENUMERATION MULTIPLICITY ESTIMATION - Test Results

Table 1: NUMBER OF SITES WHERE ALL PERSONS RESPONDED 1 TO THE USAGE QUESTION

- About 30% of shelters had all persons responding that they only used a shelter 1 night of the last 7 nights.
- Less than 5% of soup kitchens and mobile food vans had all persons responding 1.

Table 2: RESPONSES TO USAGE QUESTIONS BY TYPE OF SERVICE FACILITY

- Proportion of Usage Responses of 1 in Women's shelters and Children's Shelters 60% and 70%.
- Response rate distribution in soup kitchens and mobile food vans not as skewed.
- Imputation rates lower in shelters (14% -23%) than in either soup kitchens or mobile food vans (around 30%).

Table 3: RATIO OF LCO MULTIPLICITY ESTIMATE TO UNWEIGHTED COUNT

- The median ratio for the specified estimate is 3.83 and 25% of the LCOs have ratios greater than 4.43.
- The median ratio for Trim Option 1 is 3.28 and 25% of the LCOs have ratios greater than 3.84.
- The median ratio for Trim Option 2 is 3.18 and 25% of the LCOs have ratios greater than 3.75.
- Option 1 reduces the specified estimate by about 10% and Option 2 reduces it by about 14%.

TABLE 4: COMPARISON OF DRESS REHEARSAL RESULTS WITH CENSUS IN SAME AREA

- The multiplicity estimate for the Census was about 5 times higher than the Dress Rehearsal in Columbia and 2 times higher in Sacramento
- The Census enumeration was at the end of the month and the Dress Rehearsal enumeration at the beginning of the month. This may explain some of the difference
- The ratios of the multiplicity estimate to the unweighted count are about 1.75 to 2 times higher in the Census
- Response rates to the usage questions are about the same for shelters and higher in the Census for soup kitchens

TABLE 1: NUMBER OF SITES WHERE ALL PERSONS RESPONDED "1" TO THE USAGE QUESTION - Test Results

Code	Type of Service	Number of Sites	Number for which all responded 1	Percent of Sites
701	Regular Shelter	5414	1379	25.5%
702	~~Women's Shelter~~	462	141	30.5%
703	~~Children's Shelter~~	971	333	34.3%
704	Soup Kitchen	1826	36	2.0%
705	Mobile Food Van	158	7	4.4%

TABLE 2: RESPONSES TO USAGE QUESTIONS BY TYPE OF SERVICE

Code	Type	Usage Responses							
		1	2	3	4	5	6	7	imputed
701	Regular Shelter	76571	2169	2826	2219	2046	1770	13304	29751
	Percent of Total	58.6%	1.7%	2.2%	1.7%	1.6%	1.4%	10.2%	22.8%
Percent of respondents		75.9%	2.1%	2.8%	2.2%	2.0%	1.8%	13.2%	
702	Children's Shelter	3353	129	132	123	105	99	664	932
	Percent of Total	60.6%	2.3%	2.4%	2.2%	1.9%	1.8%	12.0%	16.8%
Percent of respondents		72.8%	2.8%	2.9%	2.7%	2.3%	2.1%	14.4%	
703	Women's Shelter	8213	228	241	253	251	247	462	1663
	Percent of Total	71.1%	2.0%	2.1%	2.2%	2.2%	2.1%	4.0%	14.4%
Percent of respondents		83.0%	2.3%	2.4%	2.6%	2.5%	2.5%	4.7%	
704	Soup Kitchen	10350	1300	4170	2390	4125	4614	6504	14529
	Percent of Total	21.6%	2.7%	8.7%	5.0%	8.6%	9.6%	13.6%	30.3%
Percent of respondents		30.9%	3.9%	12.5%	7.1%	12.3%	13.8%	19.4%	
705	Mobile Food Van	801	49	128	83	178	199	356	918
	Percent of Total	29.5%	1.8%	4.7%	3.1%	6.6%	7.3%	13.1%	33.8%
Percent of respondents		44.6%	2.7%	7.1%	4.6%	9.9%	11.1%	19.8%	

TABLE 3: RATIO OF LCO MULTIPLICITY ESTIMATE TO UNWEIGHTED COUNT - Test Results

Statistic	Specified Estimator	Trim Option 1*	Trim Option 2*
Average	3.79	3.23	3.14
Minimum	1.00	1.00	1.00
Maximum	7.00	5.93	5.75
90th Percentile	4.99	4.26	4.13
75th Percentile	4.43	3.84	3.75
Median	3.83	3.28	3.18
25th Percentile	3.14	2.63	2.58
10th Percentile	2.53	2.09	1.89
U.S. Total Estimate	980520	880938	845791

* If 100% of persons in a service facility responded "one", all persons get a weight of 1

** If 95% of persons in a service facility responded "one", all persons get a weight of 1

TABLE 4: COMPARISON OF DRESS REHEARSAL RESULTS WITH CENSUS 2000 IN SAME AREA - Test Results

	Dress Rehearsal		Census 2000	
	Columbia	Sacramento	Columbia	Sacramento
Multiplicity Estimate	833	2370	4057	4851
Unweighted Count	379	948	882	1128
Ratio: Multi. Est. / Unweighted Count	2.2	2.5	4.6	4.3
Shelters				
# Usage Responses	151	202	312	282
# Usage Nonresponses	103	382	197	383
Response Rate	59.4%	59.4%	61.2%	61.2%
Soup Kitchens				
# Usage Responses	76	197	251	238
# Usage Nonresponses	43	258	88	121
Response Rate	63.8%	43.3%	74.0%	66.3%

ESCAP MEETING NO. 24 - 12/20/00

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 24 December 20, 2000

Prepared by: Nick Birnbaum

The twenty-fourth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on December 20, 2000 at 12:00.

The agenda for the meeting was to discuss the results of weight trimming in the A.C.E. and continue assessments of the Service Based Enumeration data.

Committee Attendees:

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
Carol Van Horn

Other Attendees:

Kenneth Prewitt
Marvin Raines
Tommy Wright
Donna Kostanich
Raj Singh
Deborah Fenstermaker
Richard Griffin
Annetta Clark Smith
Felipe Kohn
Tom Mule
Denise Smith
Roxie Jones
Nick Birnbaum
Carolee Bush
Annette Quinlan
Kathleen Styles
Maria Urrutia

I. Results of A.C.E. Weight Trimming Procedure

Howard Hogan provided some historical background on this issue by explaining how two clusters in the 1990 Post-Enumeration Survey had very large sampling weights and would have had undue influence on the dual system estimates if the weights had not been trimmed or reduced.

For Census 2000, we learned from the experience of 1990, and the A.C.E. design included the over-sampling of small block clusters to reduce potentially large weights. The two clusters trimmed in 1990 were small block clusters. Additionally, we pre-specified a weight trimming procedure to be implemented if any of the sampling weights for the clusters exceeded a certain threshold. The weight trimming is implemented to:
1) reduce the effect of outlier clusters on the post-stratum dual system estimates, and
2) protect against an undue increase in sampling variance.

At this point in the meeting, Howard introduced Tom Mule who provided the Committee with a more detailed explanation of the weight trimming process implemented for the A.C.E. It was noted that the weight trimming was implemented for four outlier clusters and that the impact on the estimates for the U.S. and Puerto Rico is very minimal. In fact, it was further noted that had the weight trimming not been pre-specified at the given thresholds, it probably would not have been carried out given that the thresholds were not greatly exceeded.

II. Additional Service Based Enumeration Data

As a follow-up to the Committee's request at the December 6, 2000 meeting, DSSD provided additional data relating to the usage questions and the multiplicity estimation results. These data were discussed, and it was agreed that these data reconfirmed the concerns regarding usage response patterns that Committee members had expressed at the earlier meeting. The Committee directed Annetta Clark Smith to obtain background information on usage patterns at shelters. Ms. Smith will proceed to search the literature on this subject.

III. Next Meeting

The agenda for the next meeting, to be held on December 27, is to examine quality indicators for the A.C.E. field activities and the quality assurance program for clerical person matching.

ESCAP MEETING NO. 25 - 12/27/00

AGENDA

There was no agenda developed or used for the December 27, 2000 meeting.

ESCAP MEETING NO. 25 - 12/27/00

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Person Interviewing

A.C.E. Person Interviewing

- Person interviews of P-sample housing units

Telephoning & Personal Visit (PV)

- Computer Assisted Personal Interviewing (CAPI) software

    - CAPI instrument designed to collect information on

      nonmovers: those who lived at the sample address at the time of the interview and on census day
      inmovers: those who moved into the sample address since census day
      outmovers: those who lived at the sample address on census day but lived elsewhere at the time of the A.C.E. person interview

- Shortened elapsed time from Census Day to the A.C.E. enumeration (early start on telephone interviews)

    - Interviewers were allowed to contact households by telephone that were not part of nonresponse follow-up (the census questionnaire had to be data captured and include a telephone number).

Interview Design

- 300,913 units
- an interviewer had six weeks to complete the interview, after that the case was sent to nonresponse conversion (NRCO)
    - 1st 3 weeks required to interview an eligible household member
    - if an interview was not completed with a household member within the 1st 3 weeks, the interviewer was permitted to obtain the interview from a nonhousehold member (proxy).

    note: All LCO's except Hialeah, Florida finished PI on schedule by 9/01/00. Hialeah: The Census Bureau decided to re-interview households that did not mail back their forms. The A.C.E. PI interviews were conducted between 8/18/00-9/11/00.

Phases	**Scheduled Dates**
• Telephone	4/24/00-6/17/00
• Personal Visit (PV)	6/18/00-9/11/00

note: During the PV phase some interviews were conducted by telephone. They were units difficult to reach in person (for example, gated communities).

• NRCO	7/27/00-9/11/00

Detail Sheet:

Field Outcome Codes:	The outcome from the CAPI interview as of interview day. The computer assigns these outcome codes.
• Complete	All information obtained for current resident.
• Partial	This is a partial interview for current resident. We have names and the answers for age, sex, group quarters and second residence questions, but the answers may be don't know or refused.
• Refusal/No knwl Resp or Language Barrier	The household respondent refused, no knowledgeable respondent or language barrier. This is a rare occurrence during the telephone phase since these cases were usually reassigned to the personal visit phase.
• Vacant Interview Day	The unit was vacant on interview day. This is a rare occurrence for the telephone phase.
• Nonexistent Interview Day	The unit was nonexistent on interview day. The unit was either demolished or did not exist as a housing unit on interview day. This includes housing units found to be a business on interview day. This is a rare occurrence for the telephone phase.

Highlights

- tables exclude Puerto Rico
- different than MIS numbers because of the exclusion of PR and the addition of QA case outcomes in the tables (MIS used the QA outcome which was always 201 instead of the QA/PI outcome)

Table 1: Distribution of Person Interviews by Telephone and Personal Visit Phases
  a. by Regional Office (n)
  b. by Regional Office (%)

Table 2: Distribution of All Interviews Conducted by Interview Week-Unweighted

Table 3: a. Distribution of Interviews Conducted During the Telephone Phase by Field Outcome Code for Interview Day and Regional Office- Unweighted

  b. Distribution of Interviews (Occupied Units) Conducted During the Telephone Phase by Field Outcome Code for Interview Day and Regional Office- Unweighted

Table 4: Distribution of Interviews Conducted During the Telephone Phase by Field Outcome Code for Interview Day and by Household Member vs. Proxy- Unweighted (Percentage of Total Telephone Workload)

Table 5: a. Distribution of Interviews Conducted during the Personal Visit Phase by Field Outcome Code for Interview Day and Regional Office- Unweighted

  b. Distribution of Interviews (Occupied Units) Conducted during the Personal Visit Phase by Field Outcome Code for Interview Day and Regional Office- Unweighted

Table 6: Distribution of Interviews Conducted During the Personal Visit Phase by Field Outcome Code for Interview Day and by Household Member vs. Proxy- Unweighted (Percentage of Total PV Workload)

Table 7: Distribution of NRCO cases by Field Outcome Code-Unweighted (Percent)

Table 8: a. 1990 PES PI Results: Initial Interviews by Outcome - Unweighted

  b. 2000 A.C.E. PI Results by Field Occupied Status-Unweighted

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Table 1: Distribution of Person Interviews by Telephone and Personal Visit Phases

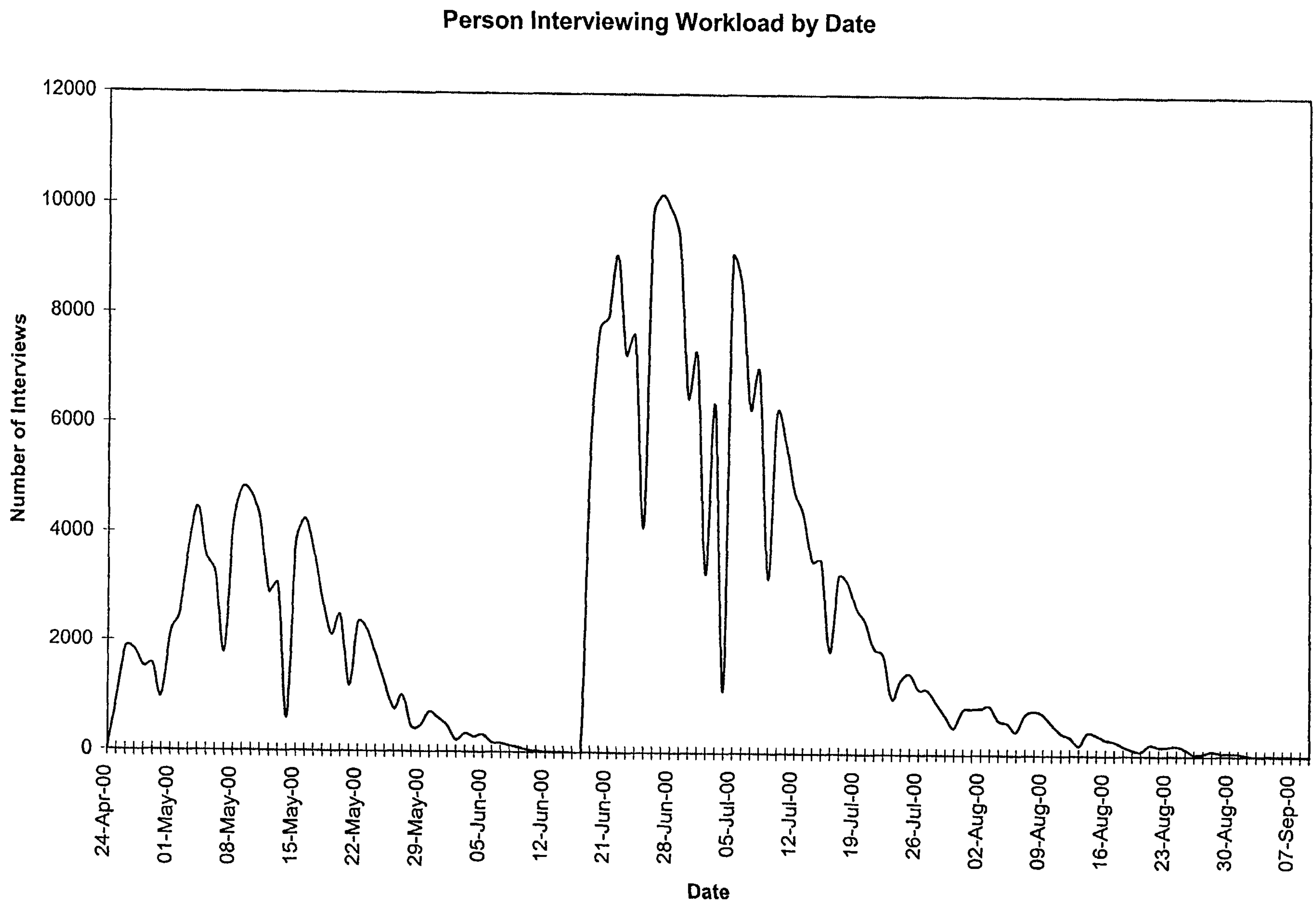
	Total Workload	Telephone Phase	PV Phase
# Interviews Conducted	300,913	88,573	212,340
Percent of PI Workload	100%	29.4	70.6

Table 1a: Distribution of Person Interviews by Telephone and Personal Visit Phases and Regional Office-Unweighted

A.C.E. Regional Office	Total Workload	Telephone Phase	PV Phase
Total	300,913	88,573	212,340
Boston	23,504	6,829	16,675
New York	17,434	3,376	14,058
Philadelphia	24,558	7,587	16,971
Detroit	23,199	7,837	15,362
Chicago	23,819	7,849	15,970
Kansas City	22,702	7,715	14,987
Seattle	24,050	7,390	16,660
Charlotte	29,027	8,077	20,950
Atlanta	27,466	8,510	18,956
Dallas	27,713	7,772	19,941
Denver	31,048	7,780	23,268
Los Angeles	26,393	7,851	18,542

Table 1b: Distribution of Person Interviews by Telephone and Personal Visit Phases and Regional Office-Unweighted, Percent

A.C.E. Regional Office	Total Workload	Telephone Phase	PV Phase
Total	100%	29.4	70.6
Boston	100%	29.1	70.9
New York	100%	19.4	80.6
Philadelphia	100%	30.9	69.1
Detroit	100%	33.8	66.2
Chicago	100%	33.0	67.0
Kansas City	100%	34.0	66.0
Seattle	100%	30.7	69.3
Charlotte	100%	27.8	72.2
Atlanta	100%	31.0	69.0
Dallas	100%	28.0	72.0
Denver	100%	25.1	74.9
Los Angeles	100%	29.7	70.3



Person Interviewing Workload by Date
Number of Interviews
12000
10000
8000
6000
4000
2000
0
24-Apr-00
01-May-00
08-May-00
15-May-00
22-May-00
29-May-00
05-Jun-00
12-Jun-00
21-Jun-00
28-Jun-00
05-Jul-00
12-Jul-00
19-Jul-00
26-Jul-00
02-Aug-00
09-Aug-00
16-Aug-00
23-Aug-00
30-Aug-00
07-Sep-00
Date

Table 2: Distribution of All Interviews Conducted by Interview Week-Unweighted

Interview Week Starting On	# Interviews Conducted Overall	Overall Percent of PI Workload
Telephone		
April 23, 2000	7,699	2.6
April 30, 2000	20,590	6.8
May 7, 2000	25,638	8.5
May 14, 2000	19,728	6.6
May 21, 2000	10,497	3.5
May 28, 2000	3,232	1.1
June 4, 2000	1,154	0.4
June 11, 2000	35	0.0
Subtotal	88,573	29.4
Personal Visit		
June 18, 2000	45,204	15.0
June 25, 2000	57,241	19.0
July 2, 2000	41,642	13.8
July 9, 2000	31,344	10.4
July 16, 2000	17,038	5.7
July 23, 2000	7,764	2.6
July 30, 2000	5,057	1.7
Aug 6, 2000	3,982	1.3
Aug 13, 2000	1,756	0.6
Aug 20, 2000	939	0.3
Aug 27, 2000	336	0.1
Sept 3, 2000	36	0.0
Sept 10, 2000	1	0.0
Subtotal	212,340	70.6
Total	300,913	100%

Table 3a: Distribution of Interviews Conducted During the Telephone Phase by Field Outcome Code for Interview Day and Regional Office - Unweighted

A.C.E. Regional Office	Total Interviews	Completed Interviews	Partial Interviews	Refusal, No knwl Resp or Language Barrier	Vacant on Interview Day	Nonexistent on Interview Day
Total	88,573	84,180	4,341	32	13	7
Boston	6,829	6,650	176	2	1	0
New York	3,376	3,117	258	1	0	0
Philadelphia	7,587	7,212	371	4	0	0
Detroit	7,837	7,553	283	1	0	0
Chicago	7,849	7,469	380	0	0	0
Kansas City	7,715	7,480	234	1	0	0
Seattle	7,390	7,003	383	2	2	0
Charlotte	8,077	7,564	494	17	0	2
Atlanta	8,510	7,985	516	1	6	2
Dallas	7,772	7,336	432	2	1	1
Denver	7,780	7,453	325	0	1	1
Los Angeles	7,851	7,358	489	1	2	1

Table 3b: Distribution of Interviews (Occupied Units) Conducted During the Telephone Phase by Field Outcome Code for Interview Day and Regional Office - Unweighted

A.C.E. Regional Office	Total Interviews at Occupied Units (n)	Completed Interviews (%)	Partial Interviews (%)	Refusal, No Knwl Resp or Language Barrier (%)
Total	88,553	95.1	4.9	0.0
Boston	6,828	97.4	2.6	0.0
New York	3,376	92.3	7.6	0.0
Philadelphia	7,587	95.1	4.9	0.0
Detroit	7,837	96.4	3.6	0.0
Chicago	7,849	95.2	4.8	0.0
Kansas City	7,715	97.0	3.0	0.0
Seattle	7,388	94.8	5.2	0.0
Charlotte	8,075	93.7	6.1	0.2
Atlanta	8,502	93.9	6.1	0.0
Dallas	7,770	94.4	5.6	0.0
Denver	7,778	95.8	4.2	0.0
Los Angeles	7,848	93.8	6.2	0.0

Table 4: Distribution of Interviews Conducted During the Telephone Phase by Field Outcome Code for Interview Day and by Household Member vs. Proxy- Unweighted (Percentage of Total Telephone Workload)

	Total Interviews	Completed Interviews	Partial Interviews	Refusals, No Knwl Resp or Language Barrier	Vacant on Interview Day	Nonexistent on Interview Day
Total	88,573 (100%)	84,180 (95.0)	4,341 (4.9)	32 (0.0)	13 (0.0)	7 (0.0)
Hhlder	88,522 (99.9)	84,179 (95.0)	4,340 (4.9)	3 (0.0)	0 (0.0)	0 (0.0)
Proxy	51 (0.1)	1 (0.0)	1 (0.0)	29 (0.0)	13 (0.0)	7 (0.0)

Table 5a: Distribution of Interviews Conducted during the Personal Visit Phase by Field Outcome Code for Interview Day and Regional Office - Unweighted

A.C.E. Regional Office	Total Interviews	Completed Interviews	Partial Interviews	Refusal, No Knwl Resp or Language Barrier	Vacant on Interview Day	Nonexistent on Interview Day
Totals	212,340	168,382	9,879	341	29,649	4,089
Boston	16,675	14,380	96	21	1,978	200
New York	14,058	11,843	1,100	20	769	326
Philadelphia	16,971	13,513	978	20	2,148	312
Detroit	15,362	12,402	633	10	2,103	214
Chicago	15,970	12,996	801	41	1,901	231
Kansas City	14,987	11,974	538	19	2,123	333
Seattle	16,660	13,621	857	37	1,853	292
Charlotte	20,950	15,496	1,050	74	3,868	462
Atlanta	18,956	13,826	980	9	3,770	371
Dallas	19,941	15,747	823	14	2,860	497
Denver	23,268	17,407	857	5	4,424	575
Los Angeles	18,542	15,177	1,166	71	1,852	276

Table 5b: Distribution of Interviews at Occupied Units Conducted during the Personal Visit Phase by Field Outcome Code for Interview Day and Regional Office - Unweighted

A.C.E. Regional Office	Total Interviews at Occupied Units (n)	Completed Interviews (%)	Partial Interviews (%)	Refusal, No Knwl Resp or Language Barrier (%)
Totals	178,602	94.3	5.5	0.2
Boston	14,497	99.2	0.7	0.1
New York	12,963	91.4	8.5	0.1
Philadelphia	14,511	93.1	6.7	0.1
Detroit	13,045	95.1	4.8	0.1
Chicago	13,838	93.9	5.8	0.3
Kansas City	12,531	95.6	4.3	0.1
Seattle	14,515	93.8	5.9	0.3
Charlotte	16,620	93.2	6.3	0.4
Atlanta	14,815	93.3	6.6	0.1
Dallas	16,584	94.9	5.0	0.1
Denver	18,269	95.3	4.7	0.0
Los Angeles	16,414	92.5	7.1	0.4

Table 6: Distribution of Interviews Conducted During the Personal Visit Phase by Field Outcome Code for Interview Day and by Household Member vs. Proxy - Unweighted (Percentage of Total PV Workload)

	Totals	Completed Interviews	Partial Interviews	Refusals, No Knwl Resp or Language Barrier	Vacant on Interview Day	Nonexistent on Interview Day
Total	212,340 (100%)	168,382 (79.3)	9,879 (4.7)	341 (0.2)	29,649 (14.0)	4,089 (1.9)
Hhlder	164,076 (77.3)	158,012 (74.4)	6,052 (2.9)	3 (0.0)	8 (0.0)	1 (0.0)
Proxy	48,264 (22.7)	10,370 (4.9)	3,827 (1.8)	338 (0.2)	29,641 (14.0)	4,088 (1.9)

Table 7: Distribution of NRCO cases by Field Outcome Code-Unweighted (Percent)

	Total NRCO Cases	NRCO Cases Completed Interviews	NRCO Cases Converted to Partial Interviews	NRCO Cases Refused	NRCO Cases Converted to Vacant	NRCO Cases Converted to Nonexistent
Total	9,735 (100%)	6,888 (70.8)	1,376 (14.1)	217 (2.2)	1,110 (11.4)	144 (1.5)
Boston	911 (100%)	562 (61.7)	32 (3.5)	7 (0.8)	275 (30.2)	35 (3.8)
New York	2,260 (100%)	1,727 (76.4)	340 (15.0)	6 (0.3)	146 (6.5)	41 (1.8)
Philadelphia	777 (100%)	543 (69.9)	122 (15.7)	12 (1.5)	90 (11.6)	10 (1.3)
Detroit	497 (100%)	345 (69.4)	109 (21.9)	0 (0.0)	41 (8.2)	2 (0.4)
Chicago	621 (100%)	433 (69.7)	94 (15.1)	35 (5.6)	45 (7.2)	14 (2.3)
Kansas City	235 (100%)	168 (71.5)	37 (15.7)	6 (2.6)	23 (9.8)	1 (0.4)
Seattle	1,501 (100%)	1,112 (74.1)	190 (12.7)	35 (2.3)	143 (9.5)	21 (1.4)
Charlotte	924 (100%)	564 (61.0)	140 (15.2)	51 (5.5)	154 (16.7)	15 (1.6)
Atlanta	653 (100%)	464 (71.1)	127 (19.4)	6 (0.9)	56 (8.6)	0 (0.0)
Dallas	857 (100%)	654 (76.3)	90 (10.5)	2 (0.2)	106 (12.4)	5 (0.6)
Denver	150 (100%)	108 (72.0)	36 (24.0)	0 (0.0)	6 (4.0)	0 (0.0)
Los Angeles	349 (100%)	208 (59.6)	59 (16.9)	57 (16.3)	25 (7.2)	0 (0.0)

1990 PES and 2000 A.C.E. Results

Introduction

In 1990 the Census Bureau conducted a Post-Enumeration Survey (PES). The 1990 PES sample consisted of 166,065 housing units. Field interviewing was completed by July 1990 in most areas and by early September for all areas. There are many differences between the 1990 and PES and 2000 A.C.E. Some of the differences are:

- The 1990 PES utilized paper questionnaires while the 2000 A.C.E. utilized CAPI for personal visit and telephone interviewing.
- The 1990 PES did not have a telephone phase as the 2000 A.C.E. did.

1990 PES Results

1990 PES results are provided in Table 8a. Note that the cases nonexistent on interview day in 1990 are excluded from these tables. **These data are not directly comparable to the 2000 results presented in Table 8b below.**

Table 8a presents the interview results by outcome as presented in *Hogan, H. " The 1990 Post-Enumeration Survey: Operations and Results", Journal of the American Statistical Association, September 1993, Vol.88, pgs. 1047-1060.*

Table 8a: 1990 PES PI Results: Initial Interviews by Outcome-Unweighted

	1990 PES	
	Number	Percent of Occupied Units
Total Housing Units	166,065	
Vacant	22,247	
Occupied	143,818	100.0
Interviews		
Household Member	134,808	93.7
Other	6,745	4.7
Noninterviews	2,265	1.6

2000 A.C.E. PI Results

Table 8b provides the results of the A.C.E. interviewing by Field Occupied Status (Results from this table are not directly comparable to the results from 1990 in Table 8a.).

Table 8b: 2000 A.C.E. PI Results by Field Occupied Status-Unweighted

	2000 ACE	
	Number	Percent of Occupied Units
Total Housing Units	300,913	
Nonexistent Housing Units	4,096	
Total Housing Units Excluding Nonexistent Units	296,817	
Vacant	29,662	
Occupied	267,155	100.0
Interviews		
Household Member	252,583	94.5
Other	14,199	5.3
Noninterviews	373	0.1

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Quality Assurance on Person Matching

Overview:

- Tri-level review system:

Person Matching Staff		
Group	# of matchers	training
Clerks	226	3 weeks BFU, 1 week AFU
Technicians	46	3 weeks BFU, 1 week AFU, plus 3 months of practice and design training
Analysts	16*	experience with previous tests and censuses; aid in development of clerk and tech training

*Including 3 headquarters employees

- Workload

Matching Work		
Stage	Workunit	Workload
BFU	Clusters	10644 clusters
AFU	Batches	4070 batches

Workload by Stage with Percent of Total		
	BFU	AFU
Clerks	10644 (100%)	4070 (100%)
Techs	5346 (50.2%)	3873 (95.16%)
Analysts	1360 (12.78%)	3363 (82.63%)

- Significant Changes
    - BFU–A code change that is determined to affect records sent to the field for person followup is significant
    - AFU–A code change that is determined to affect the match and residence status or enumeration status is significant

- Routing Clusters and Batches Through Matching:
    - Person Matching Review and Coding System (PERMaRCS) is used to count significant changes and make sampling decisions
        - BFU
            - Prequalification–Each clerk and technician was prequalified using a prespecified set of clusters.
                - Qualified–The user was set to sampling from the beginning of matching
                - Not qualified–The user had 100% of their work reviewed from the beginning of matching until they review 200 records. At this point, a decision is made (see below)
            - Decisions–Every 50 records that is reviewed by a higher level for a user, a decision is made to put the clerk in sampling or into 100% review
                - If the change rate for the clerk or tech is below 4%, the user is placed into sampling
                - If the change rate for the clerk or tech is greater than 4%, the user is placed into 100% review
            - Types of Clusters Sent to a Higher Review
                - Selected–Clusters selected for QA by sampling are completely reviewed by the next level to find errors.
                - Must-Do–Clerks and techs are allowed to flag clusters with difficult cases for further review by a more experienced matcher; the system also flags certain clusters for higher review if certain conditions are present.

- AFU
  - Worked in batches–a group of approximately 30 person followup forms to be reviewed
  - Only technicians were prequalified for AFU
  - The same decision algorithm applied as in BFU

QA Parameters				
	# Records Reviewed in 100%	Sampling Rate (1 in N)	# Records Reviewed in Sampling	Cutoff for Sampling
BFU				
Clerk	200	6	50	4%
Tech	200	10	50	4%
AFU				
Clerk	100	6	50	4%
Tech	100	10	50	4%

- Assumptions:

  - The change rate is an overestimate of the true error–In clusters not selected for QA (those clusters that are sent to a higher level of review for difficult situations), only those records that have a code at a higher level are considered to be reviewed. In reality, many more records were reviewed.

  - Analysts have no error–Due to their extensive training and specific knowledge of the task, analysts are assumed to have no errors

Modeling the Change Rate

We used three different models to estimate the overall change rates in BFU for both clerks and technicians. For any given user, we classified records four ways:

- Randomly sampled for review (X)
- Not sampled for higher review, but part of a cluster that a higher level user worked, and the higher level user coded (Y')
- Not sampled for higher review, but part of a cluster that a higher level user worked, and a higher level user did not code(Y")
- Not sampled for a higher review and not reviewed by a higher level (Z)

From the records in X, we had individual change rates, generalized here as p_x, for a given user. The sum of Y' and Y" was Y, a cluster or batch that the system did not sample for higher review, but that a higher level matcher worked. Using the proportion p_x, we estimated the overall change rate (Equations 4 and 6) and outgoing quality (Equations 5 and 7) for the remaining records. For the clerk level, records were considered part of Y if a technician reviewed the cluster or batch, but the workunit was not sampled for QA. For the technician level, records were considered part of Y if an analyst reviewed the cluster or batch, but the workunit was not sampled for QA.

Model	Y (clusters/batches reviewed, not selected for QA)	Y' (records in clusters/batches not selected for QA, but coded)	Estimation Formula
1	Random	Random	$\frac{p_x*(Y''+Z)}{X+Y+Z}$
2	Random	Not Random	$\frac{p_x*(Z)}{X+Y+Z}$
3	Not Random	n/a	$\frac{[p_x*(Y+Z)]-Y'}{X+Y+Z}$

Table 21: Overall Change Rate and Outgoing Quality Rate by Stage

Stage	Model 1		Model 2		Model 3	
	Overall Change Rate	Outgoing Quality Rate	Overall Change Rate	Outgoing Quality Rate	Overall Change Rate	Outgoing Quality Rate
BFU Clerk	0.59%	99.41	0.52%	99.48	0.44%	99.56
BFU Technician	0.23%	99.77	0.22%	99.78	0.20%	99.80
BFU Analyst	0.00%	100	0.00%	100	0.00%	100
AFU Clerk	0.95%	99.05	0.11%	99.89	0.30%	99.70
AFU Technician	0.71%	99.29	0.13%	99.87	0.24%	99.76
AFU Analyst	0.00%	100	0.00%	100	0.00%	100

ESCAP MEETING NO. 25 - 12/27/00

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting #25

December 27, 2000

Prepared by: Nick Birnbaum.

The twenty-fifth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on December 27, 2000 at 10:30.

The agenda for the meeting was to examine some measures of the quality of the A.C.E. data collection activities. There was also a discussion of the quality assurance (QA) program for clerical person matching.

Committee Attendees:

William Barron	Paula Schneider
Nancy Potok	Nancy Gordon
Cynthia Clark	John Thompson
Jay Waite	Bob Fay
Howard Hogan	John Long
Carol VanHorn	

Other Attendees:

Kenneth Prewitt	Marvin Raines
Donna Kostanich	Tommy Wright
Dan Childers	Tammy Adams
Debbie Fenstermaker	Nick Birnbaum
Kathleen Styles	Carolee Bush

I. Quality Measures of A.C.E. Field Activities

DSSD staff provided the Committee with background information regarding A.C.E. Person Interviewing and then discussed the actual field-generated data from the telephone and personal visit interview operations. The data presented preliminary operational measures of the outcome of field interviews, including completed and partial interview rates and refusal rates. The data were discussed and preliminary indications were that the operations had proceeded as expected and the level of missing data was comparable to 1990.

II. Clerical Person Matching

DSSD staff then discussed materials describing the Clerical Person Matching. The dependant verification system of review by clerks, technicians, and analysts was explained to the Committee. While the rules for determining matches are consistent at the different levels of review, the system was designed so that a more conservative approach was utilized at the lowest level of review (clerks), while analysts (who have significant experience with matching) were allowed to utilize a more liberal application of the rules. That is, the system was designed so that more matches are identified as you move up the hierarchy from clerk to technician to analyst.

The preliminary quality assurance (QA) results for the Clerical Person Matching operation were also briefly discussed. It was noted that this topic would need to be scheduled for additional discussion at a subsequent meeting.

III. Next Meeting

The agenda for the next meeting, scheduled for January 3, 2001, is to examine the effect of imputations on the quality of the initial census and the A.C.E.

ESCAP MEETING NO. 26 - 01/03/01

AGENDA

There was no agenda developed or used for the January 3, 2001 meeting.

ESCAP MEETING NO. 26 - 01/03/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Official Census Count EQUALS
People 'Correctly' Counted PLUS
Incorrect Records PLUS
Incomplete records PLUS
People & Records Added Late in Process
Imputed Records

True Population EQUALS
People 'Correctly' Counted PLUS
People 'Incorrectly' or 'Incompletely" Counted PLUS
People Counted Late PLUS
People Missed altogether PLUS

Net Undercount EQUALS
True Population MINUS Official Census

Or

People 'Incorrectly' or 'Incompletely" Counted PLUS
People Counted Late
People Missed altogether
MINUS
Incorrect Records PLUS
Incomplete records PLUS
People & Records Added Late in Process PLUS
Imputed Records

In general, the census is designed so that
People 'Incorrectly' or 'Incompletely" Counted PLUS
People Counted Late
EQUALS
Incorrect Records PLUS
Incomplete records PLUS
People & Records Added Late in Process PLUS
Imputed Records

So that, the net undercount equals the number of people missed altogether.

The problem for DSE is that in evaluating a census, we cannot easily differentiate between
People Missed altogether
And
People for whom there MAY be an incomplete, incorrect or imputed record or was included on a record added to the Census after ACE processing/matching.

How does this affect DSE?

1. We first try to determine the proportion of people in the POPULATION who were CORRECTLY AND COMPLETELY included in the census files AT THE TIME OF ACE MATCHING

2. We try determine the number of RECORDS included in the census files AT THE TIME OF ACE MATCHING which are complete and correct.

3. Using the results of Steps 1 & 2, we estimate the TOTAL POPULATON.

4. To compute the Net Undercount, we compare this estimate of the TOTAL POPULATION to the OFFICIAL CENSUS POPULATION, including imputations, late adds, incomplete and incorrect records, etc.

5. We adjust the census to correct for the NET UNDERCOUNT.

6. We distribute the NET UNDERCOUNT to local areas PROPORTIONALLY to the OFFICIAL CENSUS POPULATION.

Table D-1: 2000 A.C.E. Results---Post-Stratum Group

Post-Stratum Definition	M & F 0 - 17	Males 18 - 29	Females 18 - 29	Males 30 - 49	Females 30 - 49	Males 50+	Females 50+	Total
Census Counts								
Data-Defined Persons (DD)	This table	will be	repeated	for 64 post-	stratum	groups.		
Insufficient Information (II)								
Late Adds (LA)								
Total Persons (C)								
P Sample								
Nonmover Sample Size								
Inmover Sample Size								
Outmover Sample Size								
Weighted Nonmovers (N_n)								
Weighted Inmovers (N_i)								
Weighted Outmovers (N_o)								
Weighted Nonmover Matches (M_n)								
Weighted Outmover Matches (M_o)								
Weighted P-Sample Persons (N_p)								
Weighted P-Sample Matches (M)								
E Sample								
E-Sample Size								
Correct Enumeration Sample Size								
Weighted E-Sample Persons (N_e)								
Weighted Correct Enumerations (CE)								
Estimates								
Dual System Estimate (DSE)								
Standard Error (SE)								
Coefficient of Variation (CV) (%)								
Coverage Correction Factor								
Standard Error (SE)								
Coefficient of Variation (CV) (%)								
Net Undercount Percent (UC) (%)								
Standard Error								

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Table 1. Post-Stratum Groups - Percent Nondata Defined (ii)

Source: Final HCEF; Percents averaged over the 7 age/sex groups. ii component excludes late adds.

Domain	Tenure	MSA /TEA	High Return Rate				Low Return Rate			
			N	M	S	W	N	M	S	W
Domain 7 (Non-Hispanic White or Other)	Owner	Large MSA MO/MB	1.00	0.89	1.12	1.28	2.85	2.81	1.96	2.08
		Medium MSA MO/MB	0.84	0.78	1.18	1.31	2.09	1.62	1.72	2.01
		Small & NonMSA MO/MB	0.84	0.90	0.99	1.20	1.32	1.84	1.56	2.05
		Other TEA	1.60	1.10	1.17	2.86	1.64	1.50	1.57	2.65
	Non-Owner	Large MSA MO/MB	2.12				4.20			
		Medium MSA MO/MB	2.00				3.16			
		Small & NonMSA MO/MB	1.69				2.72			
		Other TEA	1.93				2.71			
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB Medium MSA MO/MB	2.40				4.40			
		Small & NonMSA MO/MB Other TEA	2.17				3.17			
	Non-Owner	Large MSA MO/MB Medium MSA MO/MB	3.41				5.55			
		Small & NonMSA MO/MB Other TEA	2.70				3.88			
Domain 3 (Hispanic)	Owner	Large MSA MO/MB Medium MSA MO/MB	2.85				4.38			
		Small & NonMSA MO/MB Other TEA	3.27				3.68			
	Non-Owner	Large MSA MO/MB Medium MSA MO/MB	3.19				4.53			
		Small & NonMSA MO/MB Other TEA	3.33				4.49			
Domain 5	Owner		3.36							
	Non-Owner		3.49							
Domain 6	Owner		2.41							
	Non-Owner		3.15							
Domain 1	Owner		4.43							
	Non-Owner		4.09							
Domain 2	Owner		2.21							
	Non-Owner		2.76							

Domain 5 = Native Hawaiian or Pacific Islander
Domain 6 = Non-Hispanic Asian
Domain 1 = American Indian or Alaska Native on Reservations
Domain 2 = American Indian or Alaska Native off Reservations
MO/MB = Mail Out/Mail Back
TEA = Type of Enumeration Area

Table 2. Post-Stratum Groups - Percent of Late Adds and Nondata Defined (LA+ii)

Source: Final HCEF; Percents are averaged over the 7 age/sex groups.

Domain	Tenure	MSA /TEA	High Return Rate N	High Return Rate M	High Return Rate S	High Return Rate W	Low Return Rate N	Low Return Rate M	Low Return Rate S	Low Return Rate W
Domain 7 (Non-Hispanic White or Other)	Owner	Large MSA MO/MB	1.53	1.12	1.34	1.56	4.92	4.15	2.29	2.58
		Medium MSA MO/MB	1.22	1.01	1.44	1.59	3.11	2.16	2.08	2.37
		Small & NonMSA MO/MB	1.25	1.17	1.28	1.53	2.09	2.27	1.99	2.51
		Other TEA	2.93	2.41	2.76	4.27	3.52	3.37	3.83	4.84
	Non-Owner	Large MSA MO/MB	2.89				5.52			
		Medium MSA MO/MB	2.68				3.89			
		Small & NonMSA MO/MB	2.41				3.51			
		Other TEA	4.57				5.76			
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB, Medium MSA MO/MB	2.80				6.21			
		Small & NonMSA MO/MB, Other TEA	3.43				5.36			
	Non-Owner	Large MSA MO/MB, Medium MSA MO/MB	4.05				6.82			
		Small & NonMSA MO/MB, Other TEA	3.79				5.37			
Domain 3 (Hispanic)	Owner	Large MSA MO/MB, Medium MSA MO/MB	3.35				5.96			
		Small & NonMSA MO/MB, Other TEA	4.33				5.91			
	Non-Owner	Large MSA MO/MB, Medium MSA MO/MB	3.86				5.83			
		Small & NonMSA MO/MB, Other TEA	4.74				6.98			
Domain 5	Owner		4.25							
	Non-Owner		4.44							
Domain 6	Owner		3.10							
	Non-Owner		3.91							
Domain 1	Owner		5.45							
	Non-Owner		5.10							
Domain 2	Owner		3.40							
	Non-Owner		3.89							

Domain 5 = Native Hawaiian or Pacific Islander
Domain 6 = Non-Hispanic Asian
Domain 1 = American Indian or Alaska Native on Reservations
Domain 2 = American Indian or Alaska Native off Reservations
MO/MB = Mail Out/Mail Back
TEA = Type of Enumeration Area

Table 3. 2000 Census Counts by A.C.E. Categories

Source: Final HCEF

	All Persons Inc/GQs	Persons Out of Scope for A.C.E.	A.C.E. Eligible Population	Late Adds In Scope for A.C.E.	iis not LAs In Scope for A.C.E.	Percent Late Adds	Percent iis not LAs
United States	281,421,906	7,834,909	273,586,997	2,315,136	5,691,184	0.85	2.08
Alabama	4,447,100	114,720	4,332,380	55,414	95,851	1.28	2.21
Alaska	626,932	75,625	551,307	6,612	7,633	1.20	1.38
Arizona	5,130,632	109,850	5,020,782	37,446	157,921	0.75	3.15
Arkansas	2,673,400	73,908	2,599,492	36,566	36,467	1.41	1.40
California	33,871,648	819,754	33,051,894	176,915	898,857	0.54	2.72
Colorado	4,301,261	102,955	4,198,306	31,369	85,598	0.75	2.04
Connecticut	3,405,565	107,939	3,297,626	27,918	50,905	0.85	1.54
Delaware	783,600	24,583	759,017	3,325	22,020	0.44	2.90
District of Columbia	572,059	35,562	536,497	718	19,991	0.13	3.73
Florida	15,982,378	388,945	15,593,433	87,053	329,533	0.56	2.11
Georgia	8,186,453	233,822	7,952,631	75,650	186,794	0.95	2.35
Hawaii	1,211,537	35,782	1,175,755	15,682	34,253	1.33	2.91
Idaho	1,293,953	31,496	1,262,457	9,254	22,993	0.73	1.82
Illinois	12,419,293	321,781	12,097,512	110,524	298,707	0.91	2.47
Indiana	6,080,485	178,154	5,902,331	29,403	129,962	0.50	2.20
Iowa	2,926,324	104,169	2,822,155	18,804	27,179	0.67	0.96
Kansas	2,688,418	81,950	2,606,468	17,666	31,971	0.68	1.23
Kentucky	4,041,769	114,804	3,926,965	50,377	45,727	1.28	1.16
Louisiana	4,468,976	135,965	4,333,011	43,319	81,724	1.00	1.89
Maine	1,274,923	34,912	1,240,011	16,683	15,601	1.35	1.26
Maryland	5,296,486	134,056	5,162,430	29,518	129,200	0.57	2.50
Massachusetts	6,349,097	221,216	6,127,881	55,733	95,497	0.91	1.56
Michigan	9,938,444	249,889	9,688,555	51,910	129,891	0.54	1.34
Minnesota	4,919,479	135,883	4,783,596	30,441	59,158	0.64	1.24
Mississippi	2,844,658	95,414	2,749,244	37,401	46,115	1.36	1.68
Missouri	5,595,211	162,058	5,433,153	48,810	65,203	0.90	1.20
Montana	902,195	24,762	877,433	12,778	13,890	1.46	1.58
Nebraska	1,711,263	50,818	1,660,445	10,204	16,504	0.61	0.99
Nevada	1,998,257	33,675	1,964,582	12,989	63,990	0.66	3.26
New Hampshire	1,235,786	35,539	1,200,247	12,837	23,845	1.07	1.99
New Jersey	8,414,350	194,821	8,219,529	70,793	165,042	0.86	2.01
New Mexico	1,819,046	36,307	1,782,739	28,146	51,606	1.58	2.89
New York	18,976,457	580,461	18,395,996	314,899	588,922	1.71	3.20
North Carolina	8,049,313	253,881	7,795,432	90,078	127,814	1.16	1.64
North Dakota	642,200	23,631	618,569	6,078	6,203	0.98	1.00
Ohio	11,353,140	299,121	11,054,019	61,602	124,151	0.56	1.12
Oklahoma	3,450,654	112,375	3,338,279	32,680	42,799	0.98	1.28
Oregon	3,421,399	77,491	3,343,908	18,363	57,204	0.55	1.71
Pennsylvania	12,281,054	433,301	11,847,753	88,634	186,434	0.75	1.57
Rhode Island	1,048,319	38,816	1,009,503	9,512	23,823	0.94	2.36
South Carolina	4,012,012	135,037	3,876,975	41,719	83,160	1.08	2.14
South Dakota	754,844	28,418	726,426	7,045	10,112	0.97	1.39
Tennessee	5,689,283	147,946	5,541,337	48,024	88,597	0.87	1.60
Texas	20,851,820	561,109	20,290,711	177,560	534,163	0.88	2.63
Utah	2,233,169	40,480	2,192,689	13,252	38,703	0.60	1.77
Vermont	608,827	20,760	588,067	9,787	13,071	1.66	2.22
Virginia	7,078,515	231,398	6,847,117	49,795	106,503	0.73	1.56
Washington	5,894,121	136,382	5,757,739	29,541	109,031	0.51	1.89
West Virginia	1,808,344	43,147	1,765,197	29,139	15,104	1.65	0.86
Wisconsin	5,363,675	155,958	5,207,717	28,776	83,348	0.55	1.60
Wyoming	493,782	14,083	479,699	6,394	12,414	1.33	2.59

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

January 3, 2001 Draft

Census 2000 Imputation of Housing Unit Status and Household Population - Results

Background

Following data collection activities, an inventory of census housing units is established. Information about the housing unit may be missing. During the "Unclassified" process, missing data is imputed (filled in) for:

- Unclassified units - Units with undetermined status of occupied, vacant, or nonexistent.

- Missing household population - Units determined to be occupied, but the number of persons living there is not known.

A subsequent Census process, substitution, fills in the missing persons for housing units imputed occupied or any housing unit with an imputed household population count. Substitution is also necessary for occupied housing units with a known population if all the persons are missing. The substitution process and results are not discussed here. Neither substitution nor this imputation procedure is applied to the group quarters population.

Methodology

In general, the nearest-neighbor hot deck method is used for filling in housing unit status and missing population count. Although group quarters are not utilized as a source for filling in missing data, they are included in order to determine the nearest-housing unit neighbor.

The hot decking is done separately for each Local Census Office (LCO). A preliminary Hundred Percent Census Unedited File (HCUF) is used to determine the records requiring imputation and to identify donor records. This file includes three types of housing unit records:

1. All housing unit records with classified status as occupied, vacant, or nonexistent (delete).

2. All housing unit records without status assigned.

3. All housing unit records identified as double deletes (kills) based on field activities.

In general, there was conflicting information about the double delete records. These were determined as nonexistent through at least two census operations. For the final HCUF, all records identified or imputed as delete are removed. Note also that the preliminary HCUF included housing unit records that were later determined to be duplicate records. These were not identified as such on the preliminary HCUF.

Three types of units requiring imputation:

- Units classified as occupied but with no population count.
- Unclassified units that we know exist. These are either occupied or vacant.
- Unclassified units for which we know nothing. These are either occupied, vacant, or delete.

For estimation purposes, six categories are defined. Each of the three types of units above are divided into two groups: single unit addresses and multi-unit addresses. Table A defines each estimation category by the type of donee record and the associated donor pool.

Table A. Estimation Categories and Donor Pools for each LCO

Estimation Category	**Donees**	**Donor Pool**
1. Single units 2. Multi-units	Occupied units with no household population.	Occupied units with a population count from an enumerator completed form. For mailback areas, restricted to completed forms that were subjected to field follow up activities.
3. Single units 4. Multi-units	Unclassified units which exist, but do not know whether occupied or vacant.	Occupied or vacant units from an enumerator completed form. For mailback areas, restricted to completed forms that were subjected to field follow up activities.
5. Single units 6. Multi-units	Unclassified units for which we know nothing.	Occupied, vacant, or delete units from an enumerator completed form. For mailback areas, restricted to completed forms that were subjected to field follow up activities.

Each potential donor record can be used as a donor only once. In general, the nearest donor record selected is from the same tract as the donee. For multi-unit structure records, the same multi-unit will be used as donors for multi-unit structure donees. If there are more donees than donors in a multi-unit structure, the nearest multi-unit structure in the same tract should be used for donors.

Table 1. Percent Distribution of Preliminary Housing Unit Records

Source: Preliminary HCUF

	Preliminary Records	Classified Records				Imputed Records			
		Total	Occupied	Vacant	Delete	Total	Occupied	Vacant	Delete
United States	122,534,761	99.61	86.95	8.41	4.25	0.39	0.22	0.13	0.04
Alabama	2,121,119	99.64	83.47	10.61	5.57	0.36	0.20	0.12	0.03
Alaska	280,243	99.27	79.79	13.80	5.67	0.73	0.38	0.27	0.09
Arizona	2,324,238	98.92	82.49	11.99	4.44	1.08	0.43	0.54	0.11
Arkansas	1,254,433	99.67	84.49	10.31	4.87	0.33	0.19	0.12	0.03
California	12,782,697	99.63	90.45	5.48	3.71	0.37	0.22	0.10	0.04
Colorado	1,913,012	99.67	87.38	7.74	4.56	0.33	0.18	0.12	0.03
Connecticut	1,454,964	99.71	90.23	5.74	3.73	0.29	0.19	0.08	0.02
Delaware	362,897	99.58	83.10	12.14	4.34	0.42	0.24	0.16	0.03
D.C.	288,644	99.65	85.94	9.10	4.61	0.35	0.22	0.11	0.02
Florida	7,694,983	99.37	82.90	12.34	4.13	0.63	0.30	0.28	0.05
Georgia	3,567,192	99.67	85.59	7.66	6.41	0.33	0.20	0.09	0.03
Hawaii	497,962	99.59	82.43	11.46	5.71	0.41	0.22	0.14	0.05
Idaho	564,579	99.41	84.10	10.15	5.17	0.59	0.32	0.18	0.08
Illinois	5,182,050	99.59	89.37	5.59	4.63	0.41	0.25	0.09	0.06
Indiana	2,684,279	99.53	87.90	7.21	4.42	0.47	0.28	0.13	0.06
Iowa	1,284,264	99.80	90.31	6.43	3.06	0.20	0.12	0.06	0.01
Kansas	1,181,388	99.77	88.64	7.84	3.28	0.23	0.14	0.08	0.02
Kentucky	1,860,882	99.61	86.79	8.54	4.28	0.39	0.23	0.10	0.05
Louisiana	2,004,110	99.76	84.09	9.50	6.17	0.24	0.15	0.07	0.02
Maine	682,186	99.24	76.75	19.22	3.28	0.76	0.34	0.40	0.02
Maryland	2,235,521	99.71	89.36	7.29	3.06	0.29	0.18	0.09	0.02
Massachusetts	2,746,158	99.69	89.80	6.42	3.47	0.31	0.19	0.10	0.03
Michigan	4,430,001	99.70	86.14	10.03	3.53	0.30	0.15	0.12	0.03
Minnesota	2,151,389	99.76	88.81	7.83	3.12	0.24	0.11	0.12	0.01
Mississippi	1,250,363	99.54	85.47	9.19	4.87	0.46	0.29	0.12	0.05
Missouri	2,567,782	99.75	86.36	9.55	3.84	0.25	0.13	0.10	0.02
Montana	436,287	99.46	82.75	12.19	4.53	0.54	0.30	0.19	0.04
Nebraska	747,189	99.77	89.63	7.49	2.66	0.23	0.13	0.08	0.01
Nevada	861,325	98.96	87.17	8.59	3.20	1.04	0.65	0.30	0.10
New Hampshire	573,597	99.32	83.67	12.36	3.30	0.68	0.35	0.30	0.03
New Jersey	3,473,550	99.76	89.20	7.03	3.54	0.24	0.14	0.08	0.02
New Mexico	837,405	99.17	81.92	12.03	5.22	0.83	0.47	0.28	0.08
New York	8,257,545	99.67	86.49	7.46	5.72	0.33	0.20	0.10	0.03
North Carolina	3,757,623	99.58	84.72	10.33	4.53	0.42	0.23	0.13	0.05
North Dakota	305,645	99.70	84.90	10.54	4.26	0.30	0.16	0.12	0.02
Ohio	4,999,567	99.77	89.60	6.70	3.47	0.23	0.15	0.06	0.02
Oklahoma	1,591,607	99.69	85.20	10.72	3.77	0.31	0.17	0.12	0.02
Oregon	1,544,180	99.43	87.02	7.62	4.79	0.57	0.35	0.14	0.08
Pennsylvania	5,541,364	99.67	87.20	8.44	4.03	0.33	0.19	0.11	0.03
Rhode Island	461,850	99.71	89.31	6.76	3.65	0.29	0.18	0.08	0.03
South Carolina	1,929,859	99.48	81.23	11.28	6.98	0.52	0.27	0.19	0.05
South Dakota	338,255	99.54	86.34	9.60	3.60	0.46	0.27	0.15	0.03
Tennessee	2,619,407	99.67	86.57	7.85	5.26	0.33	0.20	0.08	0.04
Texas	8,568,761	99.61	87.11	8.83	3.67	0.39	0.24	0.12	0.03
Utah	817,820	99.71	86.66	8.15	4.90	0.29	0.16	0.10	0.03
Vermont	314,498	98.90	77.51	16.63	4.77	1.10	0.61	0.47	0.02
Virginia	3,021,828	99.79	90.12	6.73	2.94	0.21	0.12	0.07	0.02
Washington	2,601,491	99.57	88.11	6.83	4.62	0.43	0.27	0.11	0.06
West Virginia	893,162	99.68	83.62	12.01	4.05	0.32	0.17	0.11	0.04
Wisconsin	2,435,916	99.48	86.32	9.52	3.65	0.52	0.23	0.23	0.05
Wyoming	237,694	99.07	81.71	12.42	4.93	0.93	0.58	0.32	0.03
Puerto Rico	1,505,654	99.33	84.74	10.28	4.31	0.67	0.45	0.16	0.06

Table 2. Distributions of Classified and Imputated Housing Unit Records

Source: Preliminary HCUF

	Preliminary Records	Distribution of Classifieds				Distribution of Imputes			
		Total	Occupied	Vacant	Delete	Total	Occupied	Vacant	Delete
United States	122,534,761	100.0	87.3	8.4	4 3	100.0	56 2	34.2	9.6
Alabama	2,121,119	100.0	83.8	10.6	5 6	100 0	56.0	34.5	9.6
Alaska	280,243	100.0	80 4	13.9	5 7	100.0	51.8	36.3	11.9
Arizona	2,324,238	100.0	83.4	12.1	4.5	100 0	39.5	50 0	10.5
Arkansas	1,254,433	100.0	84.8	10.3	4.9	100.0	56.8	35.6	7.6
California	12,782,697	100.0	90.8	5.5	3.7	100.0	60.8	28.5	10.6
Colorado	1,913,012	100.0	87.7	7.8	4.6	100.0	54.0	36.8	9.1
Connecticut	1,454,964	100.0	90.5	5.8	3.7	100.0	65.2	26.6	8.2
Delaware	362,897	100.0	83.5	12.2	4.4	100.0	55.9	37.8	6.3
D.C.	288,644	100.0	86.2	9.1	4.6	100.0	63.7	31.0	5.3
Florida	7,694,983	100.0	83.4	12.4	4.2	100.0	48.3	44.4	7.3
Georgia	3,567,192	100.0	85.9	7.7	6.4	100.0	61.7	28.5	9.8
Hawaii	497,962	100.0	82.8	11.5	5 7	100.0	53.8	33.3	12.9
Idaho	564,579	100.0	84.6	10.2	5.2	100.0	55.5	31.0	13.6
Illinois	5,182,050	100.0	89.7	5.6	4.6	100.0	62.4	23.0	14.5
Indiana	2,684,279	100.0	88.3	7.2	4.4	100 0	59.4	27.8	12.8
Iowa	1,284,264	100.0	90.5	6.4	3.1	100.0	60.1	32.8	7.1
Kansas	1,181,388	100.0	88.8	7.9	3.3	100.0	60.7	32.1	7 2
Kentucky	1,860,882	100.0	87.1	8.6	4.3	100.0	59.7	26.9	13.4
Louisiana	2,004,110	100.0	84.3	9.5	6.2	100.0	61.0	31.1	7.9
Maine	682,186	100.0	77.3	19.4	3.3	100.0	44.6	53.3	2.1
Maryland	2,235,521	100.0	89.6	7.3	3.1	100.0	59.9	32.1	8.0
Massachusetts	2,746,158	100.0	90.1	6.4	3.5	100.0	59.9	31.1	9.0
Michigan	4,430,001	100.0	86.4	10.1	3.5	100.0	49.3	42.0	8.7
Minnesota	2,151,389	100.0	89 0	7.8	3.1	100.0	45.0	49.6	5.4
Mississippi	1,250,363	100 0	85.9	9.2	4.9	100.0	63.2	25.4	11.4
Missouri	2,567,782	100 0	86.6	9 6	3.8	100.0	51.2	40.5	8.3
Montana	436,287	100.0	83.2	12 3	4.5	100.0	56.1	35.9	8.0
Nebraska	747,189	100 0	89.8	7.5	2.7	100.0	58.8	36 2	5.0
Nevada	861,325	100.0	88.1	8.7	3.2	100.0	62.1	28.3	9.6
New Hampshire	573,597	100.0	84.2	12.4	3.3	100.0	51.1	43.9	4.9
New Jersey	3,473,550	100.0	89.4	7.0	3 5	100.0	58.5	33.8	7.7
New Mexico	837,405	100.0	82.6	12.1	5.3	100.0	56.7	33.2	10 0
New York	8,257,545	100.0	86.8	7.5	5 7	100.0	60 0	31.4	8.6
North Carolina	3,757,623	100.0	85.1	10.4	4.6	100.0	55.8	31.9	12.3
North Dakota	305,645	100.0	85.2	10.6	4.3	100.0	53.9	40.6	5.5
Ohio	4,999,567	100.0	89.8	6.7	3.5	100.0	63.6	27.1	9.3
Oklahoma	1,591,607	100.0	85.5	10.8	3.8	100.0	54 7	38.0	7.3
Oregon	1,544,180	100 0	87.5	7.7	4.8	100 0	61.1	25.5	13.4
Pennsylvania	5,541,364	100.0	87.5	8 5	4.0	100.0	56 6	34.4	9.0
Rhode Island	461,850	100.0	89 6	6.8	3.7	100.0	62.0	27.2	10.8
South Carolina	1,929,859	100.0	81.7	11.3	7.0	100.0	52.8	36.7	10.5
South Dakota	338,255	100 0	86.7	9.6	3 6	100.0	60.3	33.7	6 0
Tennessee	2,619,407	100.0	86.9	7.9	5.3	100.0	61.8	24 8	13.4
Texas	8,568,761	100.0	87.4	8 9	3 7	100 0	61.7	30 6	7.8
Utah	817,820	100.0	86.9	8.2	4.9	100.0	53.7	34.5	11 8
Vermont	314,498	100.0	78.4	16 8	4.8	100.0	55.2	42.6	2.1
Virginia	3,021,828	100.0	90 3	6.7	2.9	100.0	57.8	32.7	9.4
Washington	2,601,491	100 0	88.5	6.9	4.6	100.0	61.4	25 6	13 0
West Virginia	893,162	100 0	83.9	12.0	4 1	100 0	53.6	34.5	12.0
Wisconsin	2,435,916	100.0	86.8	9.6	3.7	100 0	45 3	44 8	9.9
Wyoming	237,694	100.0	82.5	12.5	5.0	100.0	62.7	33.9	3.4
Puerto Rico	1,505,654	100.0	85 3	10.3	4.3	100.0	67.1	23.8	9.2

Table 3. Distributions of Occupied Housing Units

Source: Preliminary HCUF

	Occupied Housing Units	Percent of Occupied Housing Units		
		W/Pop	No Pop	Imputed
United States	106,810,995	99 6	0.2	0.3
Alabama	1,774,638	99.3	0.4	0.2
Alaska	224,680	99.4	0.1	0.5
Arizona	1,927,246	99.1	0.4	0.5
Arkansas	1,062,210	99.7	0.1	0 2
California	11,589,923	99.6	0.2	0.2
Colorado	1,674,879	99.6	0.2	0.2
Connecticut	1,315,635	99.7	0.1	0.2
Delaware	302,431	99.3	0.4	0.3
D.C.	248,701	99.1	0.7	0.3
Florida	6,402,468	99.5	0.2	0.4
Georgia	3,060,539	99 6	0.2	0.2
Hawaii	411,554	99.5	0 3	0.3
Idaho	476,620	99.5	0.1	0 4
Illinois	4,644,557	99.4	0.3	0.3
Indiana	2,366,985	99.6	0.1	0.3
Iowa	1,161,328	99.8	0 0	0.1
Kansas	1,048,870	99.7	0.1	0.2
Kentucky	1,619,379	99.7	0.1	0.3
Louisiana	1,688,110	99.7	0 2	0 2
Maine	525,860	99.5	0.1	0.4
Maryland	2,001,504	99.5	0.3	0.2
Massachusetts	2,471,089	99.6	0 2	0.2
Michigan	3,822,709	99.8	0.1	0.2
Minnesota	1,912,965	99 8	0.1	0.1
Mississippi	1,072,396	99.6	0.1	0.3
Missouri	2,220,880	99.8	0.1	0.1
Montana	362,346	99.5	0.2	0.4
Nebraska	670,662	99 8	0.0	0.1
Nevada	756,410	99.0	0.3	0.7
New Hampshire	481,900	99.3	0.3	0.4
New Jersey	3,103,098	99.6	0.2	0.2
New Mexico	689,959	98.9	0.5	0 6
New York	7,158,346	99.4	0.4	0.2
North Carolina	3,192,244	99.6	0.1	0.3
North Dakota	259,975	99.7	0.1	0 2
Ohio	4,486,700	99.8	0.0	0.2
Oklahoma	1,358,786	99.7	0 1	0.2
Oregon	1,349,157	99.4	0.2	0 4
Pennsylvania	4,842,361	99.6	0 2	0 2
Rhode Island	413,305	99 4	0 4	0 2
South Carolina	1,572,867	99.4	0 2	0.3
South Dakota	292,990	99.5	0.1	0.3
Tennessee	2,272,852	99.6	0 2	0.2
Texas	7,484,679	99.5	0.2	0.3
Utah	709,971	99.7	0.1	0.2
Vermont	245,669	99.1	0.1	0.8
Virginia	2,727,095	99.7	0.2	0.1
Washington	2,299,211	99.6	0.1	0.3
West Virginia	748,374	99.8	0 0	0 2
Wisconsin	2,108,271	99.6	0.1	0.3
Wyoming	195,611	99 0	0.3	0.7
Puerto Rico	1,282,691	99 3	0 2	0 5

Table 4. Distribution of Population in Housing Units

Source: Preliminary HCUF

	Population in Housing Units	Percent Not Imputed	Percent Imputed
United States	277,216,072	99.6	0.4
Alabama	4,431,269	99.3	0.7
Alaska	616,581	99.5	0.5
Arizona	5,089,402	99.1	0.9
Arkansas	2,650,345	99.7	0.3
California	33,305,218	99.6	0.4
Colorado	4,240,191	99.6	0.4
Connecticut	3,333,855	99.7	0.3
Delaware	768,475	99.3	0.7
D.C.	537,402	99.1	0.9
Florida	15,753,798	99.5	0.5
Georgia	8,098,137	99.6	0.4
Hawaii	1,201,748	99.4	0.6
Idaho	1,282,573	99.5	0.5
Illinois	12,242,877	99.4	0.6
Indiana	5,983,398	99.6	0.4
Iowa	2,853,073	99.8	0.2
Kansas	2,635,552	99.7	0.3
Kentucky	4,003,253	99.7	0.3
Louisiana	4,421,802	99.7	0.3
Maine	1,258,924	99.5	0.5
Maryland	5,214,617	99.5	0.5
Massachusetts	6,196,929	99.6	0.4
Michigan	9,786,649	99.8	0.2
Minnesota	4,831,290	99.8	0.2
Mississippi	2,821,549	99.6	0.4
Missouri	5,501,497	99.8	0.2
Montana	886,695	99.4	0.6
Nebraska	1,672,247	99.8	0.2
Nevada	1,978,345	99.0	1.0
New Hampshire	1,218,691	99.4	0.6
New Jersey	8,322,810	99.7	0.3
New Mexico	1,816,619	98.9	1.1
New York	18,684,235	99.4	0.6
North Carolina	7,950,518	99.6	0.4
North Dakota	626,214	99.7	0.3
Ohio	11,162,314	99.8	0.2
Oklahoma	3,381,873	99.7	0.3
Oregon	3,382,779	99.4	0.6
Pennsylvania	12,013,242	99.6	0.4
Rhode Island	1,022,301	99.4	0.6
South Carolina	3,980,553	99.4	0.6
South Dakota	733,976	99.5	0.5
Tennessee	5,644,874	99.6	0.4
Texas	20,553,586	99.5	0.5
Utah	2,219,533	99.8	0.2
Vermont	600,928	99.1	0.9
Virginia	6,919,388	99.7	0.3
Washington	5,832,383	99.6	0.4
West Virginia	1,795,948	99.8	0.2
Wisconsin	5,270,716	99.6	0.4
Wyoming	484,900	99.0	1.0
Puerto Rico	3,833,007	99.3	0.7

Table 5. Population per Housing Unit

Source. Preliminary HCUF

	Total	No Pop. Imputation	With Pop. Imputation
United States	2.60	2.60	2.56
Alabama	2.50	2.50	2.50
Alaska	2.74	2.74	2.64
Arizona	2.64	2.64	2.67
Arkansas	2.50	2.50	2.49
California	2.87	2.87	2.66
Colorado	2.53	2.53	2.49
Connecticut	2.53	2.53	2.47
Delaware	2.54	2.54	2.45
D.C.	2.16	2.16	2.11
Florida	2.46	2.46	2.40
Georgia	2.65	2.65	2.58
Hawaii	2.92	2.92	3.05
Idaho	2.69	2.69	2.65
Illinois	2.64	2.64	2.54
Indiana	2.53	2.53	2.55
Iowa	2.46	2.46	2.43
Kansas	2.51	2.51	2.62
Kentucky	2.47	2.47	2.46
Louisiana	2.62	2.62	2.58
Maine	2 39	2.39	2.40
Maryland	2.61	2.61	2.50
Massachusetts	2.51	2.51	2.38
Michigan	2.56	2.56	2.51
Minnesota	2.53	2.53	2.38
Mississippi	2 63	2.63	2 52
Missouri	2.48	2.48	2 43
Montana	2.45	2.45	2.53
Nebraska	2.49	2.49	2.41
Nevada	2.62	2.62	2.41
New Hampshire	2.53	2.53	2.36
New Jersey	2.68	2.68	2.64
New Mexico	2.63	2.63	2.72
New York	2.61	2.61	2.58
North Carolina	2.49	2.49	2.50
North Dakota	2.41	2.41	2.39
Ohio	2.49	2.49	2 40
Oklahoma	2.49	2.49	2.48
Oregon	2.51	2.51	2.52
Pennsylvania	2.48	2.48	2.44
Rhode Island	2.47	2.47	2.53
South Carolina	2.53	2.53	2.52
South Dakota	2.51	2.50	2.80
Tennessee	2.48	2.48	2.47
Texas	2.75	2.75	2.82
Utah	3.13	3.13	2.96
Vermont	2.45	2.45	2.50
Virginia	2.54	2.54	2.51
Washington	2.54	2.54	2.58
West Virginia	2.40	2.40	2.40
Wisconsin	2.50	2.50	2.47
Wyoming	2.48	2.48	2.49
Puerto Rico	2.99	2.99	2.79

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

December 12, 2000

Note to the Record

From: Susan Love

Subject: Definition of the Census 100% Data Completeness Measures

This note documents in more detail the counts displayed in the accompanying spread sheet. The spread sheet tables reflect the 1990 counts of data completeness measures provided by Don Dalzell to John Thompson in June 2000 from the 1990 HEDF. The data are based on all 1990 census person records. The 2000 counts of completeness were produced by Don Dalzell from the HCEF D-Prime, and mirror the 1990 measures to the extent possible. The 2000 person record counts do not include those group quarters persons that were removed from the HCEF D-Prime for the recent review process. We can provide a count of the person records not included, by state, if needed.

The definitions below apply to person records as they are configured and identified on the Census 2000 HCEF. Where the concepts differ from those reflected in the 1990 census measures these differences are pointed out. The definitions make use of variables that were set by the 100% edit and allocation process, both in 2000 as well as in 1990.

Persons in Substituted Households: The total number of persons in housing units for which no population data appear on the file for anyone in the household. The edit replicates the person records for these households of the same size in their entirety from a fully enumerated household nearby. This measure includes all persons in occupied housing units whose final occupied status and/or population count was determined from hot decks, from field counts, or from census response records that consist only of a status and pop count. Using the HCEF, the count of persons in substituted households is the sum of all housing unit person records for which QDDP = 2 (the 100% person data has been substituted by the edit). QDDP is set from the allocation variables for each 100% item. QDDP is set to 2 when all the edit has set all 100% allocation variables for a person record to 7. Substitution is not used in the group quarters universe.

Persons in Unclassified Households: The number of persons in substituted households that resulted from the application of hot decks to determine the housing unit's final status and/or pop count. This is a subset of the count of persons in substituted households. Using the HCEF, the count of persons in unclassified households is the sum of all housing unit person records for which QDDP = 2 and for which SFINST has a value of 3, 4, or 5. Unclassified estimation is not used in the group quarters universe.

Persons Not Data-Defined: The number of housing unit person records for which all 100% items have been either substituted, changed, or allocated by the edit. The HCEF variable QDDP is set to 2 on the person record when the100% data has been substituted. QDDP is set to 1 on the person record when the edit changes or allocates all the 100% data for a person record that is not in a substituted household (at least one person in the household is data defined). The 100%

items for 2000 are sex, age, relationship Hispanic origin, and race. Marital status was an additional 100% item in 1990. As a result, the number of items used to determine the count of 1990 person records that were not data defined was one less than the number of items used for the 2000 person record count.

The count of group quarters person records that are not data defined was based on 5 items in 1990 but on 4 for in 2000 due to the shifting of marital status to the sample universe. The 2000 HCEF variable QDDP was not set for the group quarters person records, making it necessary to determine the count of group quarters persons who were not data defined by interpreting the allocation variables for each 100% item.

Persons Partially Defined: The number of housing unit person records that have at least one but not all 100% items substituted, changed, or allocated by the edit. For 1990 this was a count of persons with more than one item but less than 6 were answered consistently; for 2000 this is a count of persons with more than one but less than 5 items. The group quarters measure for 1990 is the count of records with at least one but not 5 100% items not changed or allocated by the edit, while the 2000 measure is the count of records with at least one but not more than 4 items changed or allocated.

Persons Fully Defined: The number of housing unit person records that have none of their 100% items substituted, changed, or allocated. For 1990 this meant 6 items were answered consistently; for 2000 this means that 5 items were answered consistently. Similarly, the 1990 group quarters person fully defined count means that 5 items were answered consistently, and for 2000 that 4 were answered consistently.

COMPLETENES OF 100% DATA FOR PERSONS IN HOUSEHOLDS--1990 CENSUS VS 2000 CENSUS
Reflects all records on the HCEF D-prime

		1990 CENSUS						2000 CENSUS					
STATE CODE	STATE NAME	TOTAL POP IN HUS	PERSONS IN SUBSTITUTED HHLDS	PERSONS IN UNCLASSIFIED HHLDS	NOT DATA DEFINED 1/	PARTIALLY DEFINED 2/	FULLY DEFINED 3/	TOTAL POP IN HUS	PERSONS IN SUBSTITUTED HHLDS	PERSONS IN UNCLASSIFIED HHLDS	NOT DATA DEFINED 1/	PARTIALLY DEFINED 2/	FULLY DEFINED 3/
	TOTAL	242,012,129	0.66%	0.02%	0.79%	16.03%	83.18%	273,643,273	1.26%	0.43%	2.11%	10.90%	86.99%
01	AL	3,948,185	0.68%	0.01%	0.82%	17.98%	81.20%	4,332,380	1.61%	0.66%	2.26%	10.44%	87.30%
02	AK	529,342	0.55%	0.04%	0.64%	8.81%	90.55%	607,583	0.91%	0.54%	2.48%	8.15%	89.37%
04	AZ	3,584,545	1.04%	0.09%	1.26%	12.89%	85.84%	5,020,782	2.09%	0.88%	3.20%	13.18%	83.62%
05	AR	2,292,393	0.35%	0.01%	0.45%	11.81%	87.74%	2,599,492	0.81%	0.29%	1.45%	9.55%	89.01%
06	CA	29,008,161	0.73%	0.04%	0.98%	15.28%	83.74%	33,051,894	1.12%	0.36%	2.74%	15.88%	81.38%
08	CO	3,214,922	0.88%	0.02%	0.96%	12.50%	86.54%	4,198,306	1.33%	0.37%	2.08%	11.09%	86.83%
09	CT	3,185,949	1.22%	0.02%	1.29%	17.49%	81.22%	3,297,626	0.98%	0.31%	1.56%	9.97%	88.47%
10	DE	646,097	1.16%	0.02%	1.30%	14.86%	83.83%	759,017	2.19%	0.68%	2.92%	10.07%	87.02%
11	DC	565,183	2.44%	0.02%	2.62%	25.32%	72.06%	536,497	2.37%	0.89%	3.74%	16.15%	80.11%
12	FL	12,630,465	0.92%	0.01%	1.02%	16.70%	82.27%	15,593,433	1.42%	0.53%	2.14%	11.12%	86.74%
13	GA	6,304,583	0.63%	0.01%	0.76%	19.10%	80.14%	7,952,631	1.48%	0.41%	2.38%	11.48%	86.14%
15	HI	1,070,597	1.01%	0.09%	1.24%	13.81%	84.96%	1,175,755	1.47%	0.52%	2.98%	13.99%	83.03%
16	ID	985,259	0.40%	0.12%	0.61%	7.80%	91.59%	1,262,457	1.15%	0.52%	1.85%	8.09%	90.06%
17	IL	11,143,646	0.89%	0.01%	0.98%	17.55%	81.47%	12,097,512	1.57%	0.59%	2.50%	11.06%	86.44%
18	IN	5,382,167	0.50%	0.00%	0.57%	15.53%	83.91%	5,902,331	1.61%	0.43%	2.22%	8.15%	89.63%
19	IA	2,677,235	0.28%	0.00%	0.32%	13.63%	86.04%	2,822,155	0.54%	0.17%	0.98%	6.73%	92.29%
20	KS	2,394,809	0.52%	0.03%	0.59%	12.34%	87.06%	2,606,468	0.71%	0.26%	1.24%	8.47%	90.28%
21	KY	3,584,120	0.39%	0.01%	0.45%	17.62%	81.93%	3,926,965	0.72%	0.33%	1.19%	8.25%	90.56%
22	LA	4,107,395	0.73%	0.02%	0.89%	19.49%	79.62%	4,333,011	1.19%	0.33%	1.92%	10.43%	87.65%
23	ME	1,190,759	0.50%	0.02%	0.56%	11.55%	87.89%	1,240,011	0.78%	0.50%	1.28%	7.91%	90.81%
24	MD	4,667,612	0.79%	0.01%	0.91%	16.72%	82.37%	5,162,430	1.78%	0.45%	2.56%	10.16%	87.28%
25	MA	5,802,118	0.93%	0.01%	1.01%	16.84%	82.15%	6,127,881	0.99%	0.36%	1.57%	9.76%	88.66%
26	MI	9,083,605	0.41%	0.01%	0.47%	15.98%	83.55%	9,688,555	0.68%	0.22%	1.35%	8.45%	90.19%
27	MN	4,257,478	0.41%	0.01%	0.47%	12.82%	86.71%	4,783,596	0.67%	0.22%	1.26%	7.13%	91.61%
28	MS	2,503,499	0.54%	0.03%	0.78%	20.98%	78.24%	2,749,244	0.84%	0.39%	1.71%	11.57%	86.72%
29	MO	4,971,676	0.35%	0.01%	0.42%	15.88%	83.70%	5,433,153	0.67%	0.20%	1.22%	7.76%	91.02%
30	MT	775,318	0.37%	0.07%	0.48%	8.91%	90.61%	877,433	1.01%	0.55%	1.64%	7.55%	90.81%
31	NE	1,530,832	0.26%	0.03%	0.32%	11.84%	87.83%	1,660,445	0.52%	0.18%	1.00%	7.33%	91.67%
32	NV	1,177,633	0.47%	0.03%	0.60%	15.62%	83.78%	1,964,582	2.21%	0.95%	3.30%	13.20%	83.50%
33	NH	1,077,101	0.65%	0.02%	0.71%	12.97%	86.32%	1,200,247	1.47%	0.65%	2.04%	7.59%	90.37%
34	NJ	7,558,820	1.01%	0.01%	1.13%	17.81%	81.06%	8,219,529	1.21%	0.34%	2.03%	11.39%	86.58%
35	NM	1,486,262	1.07%	0.23%	1.27%	12.44%	86.29%	1,782,739	2.11%	1.12%	2.98%	15.05%	81.97%
36	NY	17,445,190	0.77%	0.02%	0.91%	19.69%	79.40%	18,395,996	2.23%	0.60%	3.24%	12.96%	83.80%
37	NC	6,404,167	0.48%	0.01%	0.57%	18.27%	81.16%	7,795,432	1.08%	0.42%	1.68%	10.20%	88.12%
38	ND	614,566	0.26%	0.08%	0.32%	8.84%	90.84%	618,569	0.57%	0.26%	1.03%	6.37%	92.60%
39	OH	10,585,664	0.35%	0.00%	0.40%	15.24%	84.36%	11,054,019	0.60%	0.20%	1.13%	7.58%	91.29%
40	OK	3,051,908	0.56%	0.02%	0.66%	14.27%	85.06%	3,338,279	0.73%	0.27%	1.30%	9.06%	89.64%
41	OR	2,776,116	0.35%	0.04%	0.45%	12.62%	86.93%	3,343,908	1.05%	0.54%	1.73%	9.11%	89.16%
42	PA	11,533,219	0.55%	0.01%	0.64%	17.01%	82.35%	11,847,753	1.01%	0.36%	1.59%	8.61%	89.80%
44	RI	964,869	1.79%	0.01%	1.89%	18.87%	79.24%	1,009,503	1.73%	0.63%	2.37%	9.94%	87.69%
45	SC	3,370,160	0.83%	0.02%	0.97%	18.64%	80.39%	3,876,975	1.53%	0.58%	2.21%	10.43%	87.36%
46	SD	670,163	0.53%	0.11%	0.67%	10.76%	88.58%	726,426	0.81%	0.52%	1.42%	6.87%	91.91%
47	TN	4,748,056	0.43%	0.01%	0.52%	16.71%	82.76%	5,541,337	1.08%	0.40%	1.62%	9.21%	89.17%
48	TX	16,593,063	0.73%	0.03%	0.89%	14.77%	84.33%	20,290,711	1.62%	0.51%	2.67%	13.08%	84.25%
49	UT	1,693,802	0.41%	0.11%	0.63%	11.19%	88.18%	2,192,689	0.72%	0.25%	1.78%	8.56%	89.66%
50	VT	541,116	0.54%	0.05%	0.61%	9.30%	90.09%	588,067	1.61%	0.92%	2.26%	7.67%	90.07%
51	VA	5,978,058	0.46%	0.01%	0.54%	16.32%	83.14%	6,847,117	1.00%	0.34%	1.58%	9.32%	89.10%
53	WA	4,746,161	0.47%	0.03%	0.56%	12.79%	86.65%	5,757,739	1.16%	0.43%	1.91%	10.01%	88.08%
54	WV	1,756,566	0.50%	0.01%	0.56%	16.52%	82.92%	1,765,197	0.48%	0.23%	0.87%	8.32%	90.81%
55	WI	4,758,171	0.31%	0.01%	0.39%	13.49%	86.12%	5,207,717	1.02%	0.37%	1.62%	7.37%	91.01%
56	WY	443,348	0.65%	0.11%	0.76%	8.87%	90.38%	479,699	1.85%	1.02%	2.71%	7.74%	89.55%

ESCAP MEETING NO. 26 - 01/03/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 26 January 3, 2001

Prepared by: Nick Birnbaum

The twenty-sixth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on January 3, 2001 at 10:30.

The agenda for the meeting was to discuss the impact of whole person imputations and late adds in the census on Dual System Estimation (DSE).

Committee Attendees:

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Other Attendees:

Kenneth Prewitt	Deborah Fenstermaker
Marvin Raines	Roxie Jones
Tommy Wright	Kathleen Styles
Donna Kostanich	Nick Birnbaum
Raj Singh	Carolee Bush
Gregg Robinson	Annette Quinlan
Richard Griffin	Maria Urrutia

I. Effect of Whole Person Imputations and Late Adds in the Census on DSE

Howard Hogan began his presentation by explaining the DSE methodology used for incomplete, incorrect, imputed, or records added late in the process (after A.C.E. matching). The DSE methodology is based on calculating an estimate for the total population based on the records in the census that have information suitable for matching. By comparing this estimate to the total census population (including imputations, etc.) an estimate of net undercount can be obtained. To put it another way, at the time of A.C.E. matching, we try to determine the proportion of people in the population who were completely and correctly included in the census files and the number of records included in the census files that are complete and correct. DSE uses this information to produce estimates of the true population. These estimates are compared to the full census results to produce measures of undercounts.

This method has been successfully used in past censuses. The basic assumption is that the P-sample probability of inclusion for the persons these records represent is the same as for other persons within the post-stratum. The primary effect on the DSE estimates from this procedure is related to the level of variance for the DSE.

Howard then presented data, attachment, showing the percent nondata defined (insufficient information or duplicates) by post-stratum group. It was noted that the percentages varied considerably by mail return, tenure, and other post-stratum variables. Next, Howard presented data showing the percent of late adds and nondata defined by post-stratum group. It was noted that these data also varied considerably. This raises concerns about the synthetic assumption. It was further noted that the level of imputations and residual duplicates in the census will have a non-trivial effect on the variance of the estimates.

Howard then presented data, attached, showing the numbers and percentages of late adds and insufficient information cases by state. The Committee noted additional clustering in the data, again raising concerns with the synthetic assumption.

After briefly discussing the methodology, Howard then presented preliminary data on the Census 2000 imputation of housing unit status and household population. The data presented included distributions by state showing the effect of the operation on the preliminary housing unit and population counts. The purpose of these data was to identify any potential clustering effects in the levels of imputation. No outliers were identified.

Finally, Howard presented final results for “completeness measures” for persons in households for Census 2000. These measures are based on, for the nation and state, the numbers and percentages of persons in substituted households, the numbers and percentages of persons in unclassified households, and the numbers and percentages of persons not data-defined, partially data-defined, and fully data-defined. The comparable data for 1990 were also presented. It was clearly noted that, both in absolute terms and proportionately, there were more whole person imputations in 2000 than in 1990. This result is in part due to less personal visit followup in 2000, compared to 1990. As mentioned earlier, the large number of not data-defined people will contribute to sampling variance in the A.C.E. estimates and also indicate data quality issues for the census.

II. Next Meeting

The agenda for the next meeting, to be held on January 10, is to discuss total error model components and how they are used in Loss Function analysis.

ESCAP MEETING NO. 27 - 01/10/01

AGENDA

There was no agenda developed or used for the January 10, 2001 meeting.

ESCAP MEETING NO. 27 - 01/10/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

OVERVIEW OF TOTAL ERROR IN THE 2000 A.C.E.

Mary H. Mulry
January 10, 2001

Background

- The decision on whether to release the adjusted 2000 Census numbers for redistricting by the April 1 deadline will use a loss function analysis as a tool for comparing the quality of the original enumeration and the A.C.E. estimates.
- Total error analysis is a fundamental building block for the loss function analysis.
- Total error analysis was used to evaluate the 1990 PES for the census adjustment decision in 1991 and for the 1992 CAPE adjustment decision concerning census adjustment for the postcensal estimates.

Purpose of total error analysis

- Total error estimation synthesizes all the sampling and nonsampling components of error in the A.C.E. to produce the net effect of these errors on the DSE.
- Total error analysis produces an estimation methodology for the loss function analysis to use in the comparison of the original enumeration and the estimates from the A.C.E.

Estimation Strategy

- First, we estimate component errors and their variances for groups of A.C.E. poststrata called evaluation poststrata.
- Then we derive estimates of component errors for each A.C.E. postratum based on the component errors for its evaluation poststrata.
- We use simulation methodology to assess the net effect of all the component errors combined and for use in the loss function analysis.

Comments and issues about measurement in 2000

- When the 1990 data is used, 1990 PES poststrata are mapped to the 2000 A.C.E. poststrata
- Component errors based on 1990 data are computed using the 1990 computer programs with minor adaptations for 2000.
- The adjustments for the component errors based on 1990 are basically to eliminate any respondents in group quarters and to scale the estimates to the 2000 population in housing units.
- The question about synthetic error is whether leaving it out of the total error model causes the loss function analysis to favor adjustment or to favor non adjustment or to remain neutral. The answer to the question depends on the covariance across areas between the effect of synthetic error and the estimate of undercount. The analyses currently under consideration will provide information about the covariance.
- Correlation bias is measured at the national level but there is a question about whether the 1990 method of distributing it to lower levels is appropriate for 2000.

Table 2. Sources of Data for Estimation of Components of Error

Error Components	Measurement in 1990	Measurement in 2000
P-sample matching error	1990 Matching Error Study	1990 Matching Error Study with adjustments for 2000
P-sample data collection error	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
P-sample fabrication	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
E-sample data collection error	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
E-sample processing error	1990 Matching Error Study	1990 Matching Error Study with adjustments for 2000
Correlation bias	1990 Demographic Analysis	2000 Demographic Analysis
Ratio estimator bias	1990 PES	2000 A.C.E.
Sampling error	1990 PES	2000 A.C.E.
Imputation error	1990 Reasonable Alternatives Imputation Study	1990 Reasonable Alternatives with adjustments for 2000
Excluded Census Data Error	1990 Excluded Data Study	Not available
Contamination of P sample by enumeration or vice versa	Shown to be negligible	Not available in time for analysis for decision
Misclassification error of records into poststrata from inconsistent reporting	Not measured	Not available in time for analysis for decision
Synthetic error	Artificial population analysis and not integrated in total error model	Under development but will not be integrated in total error model

ESCAP MEETING NO. 27 - 01/10/01

MINUTES

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 27

January 10, 2001

Prepared by: Annette Quinlan

The twenty-seventh meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on January 10, 2001 at 10:30. The agenda for the meeting was to discuss total error analysis components and how they are used in Loss Function analysis for the A.C.E. In addition, Howard Hogan began the meeting by addressing two issues with A.C.E. Estimation.

Committee Attendees:

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
Carol Van Horn

Other Attendees:

Kenneth Prewitt	Alfredo Navarro
Marvin Raines	Gregg Robinson
Tommy Wright	Nick Birnbaum
Donna Kostanich	Carolee Bush
Fay Nash	Kathleen Styles
Raj Singh	Sarah Brady
Mary Mulry	Annette Quinlan

I. A.C.E. Estimation Issues

Howard Hogan addressed two situations encountered in the missing data and post-stratification processes that had not been pre-specified. The first issue concerns the use of variances as a criterion in determining the collapsing of post-strata. The second deals with the donor pool that is used for imputing missing data for unresolved cases.

At the time of specifying the collapsing scheme for post-strata, it was believed that variance estimates would not be available for use in determining which post-strata to collapse. As a result, the specifications for collapsing post-strata did not include variances as a criterion. DSSD has produced preliminary variances for the post-strata. In reviewing these variances, it was found that the variance for one particular post-stratum indicated that collapsing would be beneficial. It was decided, after conferring with the appropriate senior staff, that this post-stratum would be collapsed with others to reduce the variance. The specifications will be modified to incorporate variance as a collapsing variable.

The second situation occurred during the imputation for missing data. At the time of developing the imputation methodology, the data from the Person Followup forms were not expected to be available for review before the missing data imputation was completed. However, the Person Followup forms were keyed and reviewed before the completion of missing data imputation. As a result of this review, it was discovered that additional information could be used to improve the imputation for two categories of E-sample persons. These categories are (1) a person who was potentially fictitious and (2) a person who was said to be living elsewhere on Census Day. The new information was therefore used to enhance the imputation rates for E-sample persons in the above categories.

II. Loss Functions

Mary Mulry presented an overview of the total error analysis for the A.C.E. and the components of this error. The handouts are attached. The total error analysis is used in building the loss function analysis through simulation methodology which synthesizes the components of error to give the net effect of the errors on the adjusted and unadjusted results. The loss functions give measures of the difference that the adjusted and unadjusted results have from an estimated "true" count obtained from the total error model. The set of data with the smaller difference (or loss) is the more accurate. One limitation of this analysis is that the "truth" is not known, but must be estimated. This estimate of the "true" count is therefore subject to uncertainty and has a certain amount of bias and variability associated with it. A simulation model is used to determine a range for this uncertainty.

As discussed at the October 25, 2000 ESCAP meeting, the input to the loss function analysis is based on a total error model used to estimate the net effect of sampling and non-sampling error in the initial census and the A.C.E. The components of the total error model are derived from the Census Bureau's evaluation studies providing various measures of sampling and non-sampling error.

The components of the total error model were discussed. All evaluation work necessary to provide estimates for each component of the total error model will not be available in time for the Committee's recommendation on adjustment. As a result, some of the components of the total error model will use 1990 data from the Matching Error Study, the Evaluation Followup, the Reasonable Alternatives Imputation Study, and the Excluded Data Study with scaling adjustments made for the 2000 A.C.E. design; these adjustments will only address factors relating to scope and population size. (For example, the A.C.E. does not include the group quarters population, whereas the
1990 PES did include components of this population.) The components of error for which 2000 data will be available are correlation bias, ratio estimator bias, and sampling error.

III. Next Meeting

The next meeting scheduled for Wednesday January 17, 2001 will discuss A.C.E. missing data results.

ESCAP MEETING NO. 28 - 01/17/01

AGENDA

Kathleen P Porter
01/16/2001 03:14 PM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, Jennifer Wight/DIR/HQ/BOC@BOC

cc:

Subject: 1/17 ESCAP Agenda

The agenda for the January 17 ESCAP Meeting scheduled from 12:00-1:30 in Rm. 2412/3 is as follows:

1. Missing Data Results - Pat Cantwell

2. Synthetic Error - Donna Kostanich and staff (time permitting)

ESCAP MEETING NO. 28 - 01/17/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

Draft

DRAFT
January 16, 2001

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-7

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Donna Kostanich
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: Patrick J. Cantwell
Statistical Communications

Subject: Accuracy and Coverage Evaluation Survey:
Missing Data Results

The attached document is a prototype of the report that we will prepare, per your request, following completion of applicable Accuracy and Coverage Evaluation Survey (A.C.E.) operations. The completed report is intended to aid the Executive Steering Committee on A.C.E. Policy (ESCAP) in its recommendation regarding the release of the statistically corrected data or the data without statistical correction as the P.L. 94-171 data. This report, together with other reports, will assess the operations and results of both the initial Census and the A.C.E. Both sets of assessments will be available to the ESCAP to aid the Committee in reaching its recommendation regarding the use of the statistically corrected data.

The attached prototype contains both empty table shells and a description of textual analysis that will assess specific aspects of the applicable operations. This report focuses on the results of the missing data procedures.

It is important to note that the conduct of the operations may lead us to modify the attached format by including additional information. It is also likely that descriptions and definitions will be enhanced or the data items could undergo revision. Conversely, we may conclude, for a variety of reasons, that some of the information set forth in the attached prototype may not be available. The attached document sets forth our conclusions prior to completion of the A.C.E. about what information would properly inform the ESCAP on this subject, but is subject to modification.

Table of Contents

List of Tables

Appendix

Accuracy and Coverage Evaluation 2000: Missing Data Results

prepared by Patrick J. Cantwell

Introduction

The Accuracy and Coverage Evaluation (A.C.E.) uses dual system estimation (DSE) to determine population estimates. The Census Bureau obtains a roster of the A.C.E. sample blocks independently of the Census. People in these blocks are interviewed and asked who lives there at the time of the interview, and who lived there on Census Day. Information is gathered to identify people who have moved in or out of the residence since the time of the census. The independent roster (P Sample) and the Census roster (E Sample) are matched; the results of the matching are then used to estimate the number of people missed by both rosters.

Estimates are calculated separately within estimation domains called post-strata. Post-stratum estimates are then used to determine coverage correction factors to be applied to all people enumerated in the Census according to their specific post-stratum. Finally, corrected counts for any geographic area will be calculated by summing the corrected counts of people in the area. An appropriate rounding method is applied to produce integer counts of people at all levels.

For each component of the dual system estimator, certain required data are not collected on some cases in the A.C.E. To address this problem we apply missing data procedures. A summary of the procedures used in the 2000 A.C.E. is given in this document; greater detail can be found in DSSD Memorandum #Q-19; specifications for programming the same procedures are provided in DSSD Memorandum #Q-25. For an overview of the changes in these procedures during the 1990s–including the 1990 Post Enumeration Survey, the Census tests of 1995 and 1996, and the Census 2000 Dress Rehearsal (conducted in 1998)–see DSSD Memorandum #Q-3.

Background

Before dual system estimates are calculated, we must account for missing information from the interviews of P-Sample people and from the matching operations. It should be noted that the term "missing data" applies after all follow-up attempts have been made. There are three main types of missing data in the A.C.E. and three processes used to correct for them.

The first type is household-level noninterviews in the A.C.E. The majority of these are households that were not interviewed because they could not be contacted or because the interview was refused. Another important component are occupied households where no one has sufficient information for matching (that is, a valid name and at least two other characteristics). The latter are also treated as household noninterviews. In general, the noninterview adjustment spreads the weights of household noninterviews among households that were interviewed in the same block cluster and type of basic address (defined below).

The second type is missing demographic characteristics used to assign people to a post-stratum. This situation occurs when a person is missing age, sex, tenure, race, or Hispanic origin. We impute tenure using a hot-deck procedure. Other characteristics, such as age, are imputed based on available demographic distributions. Still others use a combination.

Third, for a small number of A.C.E. people with "unresolved status," we assign a probability for the appropriate status. For some respondents in the P Sample, there is not enough information available to determine the match status (whether or not the person matches to someone in the E Sample in the same block cluster or the extended search area) or the resident status (whether or not the person was living in the block cluster or the associated extended search area on Census Day). Determining resident status is important for P-Sample people because Census Day residents are used to estimate the number of matches in the P Sample. Similarly, for people in the E Sample, there may not be enough information to determine whether the person was correctly enumerated. Such cases where status cannot be determined are said to be "unresolved." Generally for cases with missing status a probability is assigned based on information available about the specific case and about cases with similar characteristics.

(Note: E-Sample people without sufficient information for matching are not unresolved, but are considered as erroneous enumerations, that is, they are assigned a probability of enumeration of 0. In the P Sample, if the entire housing unit contains people without sufficient information for matching, the housing unit is treated as a noninterview (see above); otherwise, each such person has unresolved resident and match status.)

Historical Treatment of Missing Data

In the 1990 Census, the Post Enumeration Survey (PES) was conducted to measure the undercoverage in the census counts and to provide adjusted counts for various demographic groups and levels of geography. Many design features of the PES were similar to those of the A.C.E. The PES also suffered from missing data in terms of (1) noninterviewed housing units in the P Sample, (2) missing characteristics needed to assign P- and E-Sample people to post-strata, and (3) unresolved status for some cases–match status in the P Sample or enumeration status in the E Sample. The status in (3) was necessary to compute the various components of the dual system estimates (DSEs) that are the basis of the correction of census data.

For the first two types of missing data, statistical procedures similar to those currently used in the A.C.E. were applied. A noninterview adjustment was used to spread the weights of noninterviewed housing units over the set of interviewed housing units. For missing demographic characteristics, a hot-deck imputation procedure or imputation based on available distributions of characteristics was used to make sure that all PES people could be assigned to a post-stratum. These are standard practices in the field of survey methodology. The main differences between the missing data procedures for the 1990 PES and 2000 A.C.E. lie with the people who have an unresolved status.

For the PES and the A.C.E., each person in the P Sample has a probability of matching to a person in the E Sample. This probability is said to be 1 if the person matches, and 0 if the person does not match. People whose match status is "unresolved"–still unknown or unclear after all follow-up operations–must be assigned a match probability between 0 and 1 to compute the appropriate component of the DSE. Analogous situations describe resident status for P-Sample people (in the 2000 A.C.E.) and enumeration status for E-Sample people (in the PES and the A.C.E.).

In the 1990 PES, match status in the P Sample and enumeration status in the E Sample had to be determined or estimated, but resident status was not an issue. Under Mover Procedure B used in the PES, we estimated the number of movers by counting in-movers in the PES block clusters, and we estimated the match rate among movers by trying to match the P-Sample in-movers to their census records (typically, in other block clusters not in the sample). The 2000 A.C.E. uses Mover Procedure C, whereby we continue to use the in-movers as a measure of the number of movers, but we estimate the match rate among movers as that among the out-movers from the A.C.E. block clusters. Because of the difference between mover procedures, the 2000 A.C.E. has to determine census-day resident status for the P Sample–to determine who is eligible for inclusion in the DSE–while the 1990 PES did not.

The procedure for assigning probabilities to unresolved cases (for match status in the P Sample or enumeration status in the E Sample) was also different in 1990. In the PES a hierarchical logistic regression model was used to estimate the missing probabilities for unresolved cases. A very large number of demographic and geographic characteristics were used as input into the model. In contrast, under the 2000 A.C.E. procedure, all resolved and unresolved cases are separated into groups called imputation cells according to a different set of operational and demographic characteristics. Within any cell, the weighted proportion of matches (or residents, or correct enumerations) among the resolved cases is assigned as the probability of a match to all unresolved cases in that cell.

After the 1990 Census and PES, three evaluations were conducted to assess the effect of the missing data procedures on the PES estimates, P1, P2 and P3. The findings are documented in the 1990 Coverage Studies and Evaluation Memorandum Series, #A-9, #B-4 and #C-2, respectively. For these evaluations, a stratified systematic subsample of 920 PES sample block clusters was selected. Following is a description of the three evaluations.

Evaluation P1: Analysis of reasonable alternatives

Evaluation P1 focused on P-Sample match status and E-Sample enumeration status. Match and correct enumeration probabilities were imputed under several alternatives. Undercount rates for each method were computed to determine the sensitivity of the estimates to the imputation method. Results from this evaluation showed that undercount estimates were robust under reasonable imputation methods.

Evaluation P2: Distribution of missing data rates

One objective of PES evaluation was to determine the level and distribution of missing data by demographic and geographic groups, and to compare the distributions with the distribution of census undercount. Evaluation P2 examined the percent of noninterviews and proxy interviews, item imputation rates, and undercount (or overcount) estimates. Results from this evaluation showed that the rate of imputation for characteristics in the PES E Sample was higher than that in the P Sample for all characteristics. The evaluation also showed that imputation rates were highly correlated with estimated Census undercount.

Draft

Evaluation P3: Evaluation of imputation methodology for unresolved match status cases

In Evaluation P3, cases with critical missing data (noninterview, missing match status, or missing enumeration status) were re-interviewed to assess the adequacy of the missing data models for PES production. P3 compared the total number of matches and correct enumerations from re-interviewed cases with the resulting estimated numbers of matches and correct enumerations, respectively, from the corresponding PES unresolved people. The results showed a correlation between imputed match probabilities from the PES and match codes from the evaluation re-interviews. However, higher correct enumeration probabilities were not correlated with correct enumerations in the evaluation follow-up cases. For this reason, the imputation model worked better for P-Sample cases than for E-Sample cases.

For A.C.E. 2000

Results from Census 2000 evaluations of missing data operations, conducted by the Planning, Research, and Evaluations Division of the Census Bureau, will not be finished in time to include in this document. Therefore, comparisons between 2000 A.C.E. and 1990 PES missing data results in this document focus on interview rates, rates and patterns of missing characteristics, and observations for unresolved cases–including rates within imputation cells and other categories of interest, and the results of the assignment of probabilities for these unresolved cases. Where direct comparisons between 2000 and 1990 are possible, we have placed the 1990 table after the Census 2000 table in this document.

Assessment

[After completing the tables and analyzing their contents, we will summarize our observations and results in this section, including our overall assessment of the level and effect of missing data on the A.C.E. survey.]

[Observations and analyses on individual tables will follow many of these tables in the subsequent pages.]

Results and Tables

A Note on Weighting

In the tables that follow, some numbers are unweighted, while others are weighted. For weighted frequencies and rates, we determine the weights as described below.

Some of the tables are listed more than once because they fall into more than one category.

Unweighted Tables: Tables 1a, 1b, 1c, 2, 3a, 3c, 3e, 5a, 6a, 6c, 7a, 7c, 11, 12; A-2, A-3a, A-4a, A-5a, A-6a, A-7a, A-8a, A-12 through A-14.

Weighted Tables with Housing Units: To produce weighted counts of interviews, noninterviews, etc., we use as housing-unit weights the initial P-Sample weights reflecting (1) the probability of selection at all stages of sampling (including the subsegmenting and sampling within large blocks) except for TES sampling, and (2) any potential trimming of the weights. Tables 1a, 1b, 1c; A-1a, A-1b, A-27, A-28.

Weighted Tables with P-Sample People: To produce measures involving P-Sample people, we use two different sets of weights as follows:

(a) For tables that measure missing rates for the characteristics age, sex, tenure, race, and Hispanic origin (listed below), we use the final P-Sample weights reflecting (1) the probability of selection at all stages of sampling including TES sampling, (2) the noninterview adjustment (a housing-unit factor applied to the people in the housing unit), and (3) any potential trimming of the weights. Tables 3b, 3d, 3e, 4a, 4b; A-2, A-3b, A-4b, A-5b, A-6b, A-7b, A-8b.

(b) For tables that summarize results related to missing resident or match status (except for Table 3c), the weights incorporate (1) and (3) in (a). That is, we use the same weights as in (a) except that the noninterview adjustment factor is not applied because it is not used to compute the probabilities assigned to unresolved cases within an imputation cell. For these tables, the weighted rates in the tables should be consistent with the weighted numbers of resolved cases. Tables 5b, 6b, 6d, 8, 9; A-9, A-10, A-15 through A-22, A-29 through A-39.

Weighted Tables with E-Sample People. To produce measures involving E-Sample people, we use the final E-Sample weights reflecting (1) the probability of selection at all stages of sampling including TES sampling, and (2) any potential trimming of the weights. Note that there is no noninterview adjustment factor for the E Sample. Tables 3b, 3d, 4a, 4b, 7b, 7d, 10; A-11, A-23 through A-26, A-29 through A-39.

Noninterview Adjustment

Noninterview adjustment is performed only on the P Sample. A.C.E. questions are asked to determine who currently lives in the household and who lived in the household on Census Day. Thus two rosters are created for each household, the Census Day roster and the A.C.E. Interview Day roster. Because of the use of Mover Procedure C estimation, there are two noninterview adjustments–one based on housing-unit status as of Census Day (i.e., the Census Day roster), and the other based on housing-unit status as of the day of the A.C.E. interview (i.e., the A.C.E. Interview Day roster).

Each of the two noninterview adjustments generally spreads the weights of noninterviewed units over interviewed units in the same noninterview adjustment cell, defined as the block cluster crossed with the type of basic address. For purposes of this adjustment, the type of basic address is grouped by single-family units, apartments, and all others.

The Census Day housing-unit status for P-Sample units is used to compute the Census-Day noninterview adjustment, which is then applied (at the appropriate level) to the person weights of non-movers and out-movers. Similarly, A.C.E. Interview Day housing-unit status is used to compute the A.C.E. Interview Day noninterview adjustment, which is then applied to the person weights of in-movers. More information can be found in DSSD Census 2000 Procedures and Operations Memorandum Series #Q-19 and #Q-25.

Results - Noninterview Adjustment

Tables 1a, 1b, 1c, 2; A-1a, A-1b, A-27, A-28.

Interview status categories:

> Interview: A unit is an interview (for the given reference date) if there is at least one person (with name and at least two demographic characteristics) who possibly or definitely was a resident of the housing unit on the given reference date.
>
> Noninterview: An occupied housing unit (as of the given reference date) that is not an interview is a noninterview.
>
> Vacant: A housing unit is vacant if no one is living in it at the time of enumeration, unless the occupants are only temporarily absent. Units temporarily occupied at the time of enumeration entirely by individuals who have a usual residence elsewhere are classified as vacant. Transient quarters, such as hotels, are housing units only if occupied. (Thus, there are no vacant housing units at hotels and the like.) New units not yet occupied are classified as vacant housing units if construction has reached a point where all exterior windows and doors are installed and final usable floors are in place. Vacant units are

excluded from the housing unit inventory if they are open to the elements. Also excluded from the housing unit inventory are units with a posted condemnation sign or units that are used entirely for nonresidential purposes.

Delete: This category is for an address that no longer qualifies as a living quarters.

Table 1a shows the unweighted and weighted A.C.E. household interview status for Census Day for the United States. Table 1b shows the interview status for the A.C.E. interview day. These tables include data for Census 2000.

For comparison, Table 1c shows the interview status for interview day in the 1990 PES. All three tables show the Total (number of housing units in the P Sample), the number of P-Sample housing units in each of the four interview status categories (Interviews, Noninterviews, Vacants, and Deletes) and the Interview Rate.

Definition of interview rate: The unweighted (weighted) interview rate is the unweighted (weighted) number of interviews divided by the unweighted (weighted) sum of interviews and noninterviews. See Tables A-1a and A-1b in the Appendix for weighted A.C.E. interview status for census day and A.C.E. interview day respectively for each state.

These tables summarize components of A.C.E. interview rates and allow comparison between Census day and A.C.E. interview day.

Table 1a. Status of A.C.E. Household Interviews[1] for Census Day: 2000 Data

	Number (unweighted)	Number (weighted)	Percent (weighted)
U.S. Total (Housing Units)	300,913	115,650,208	100.0%
Interviews	254,175	99,166,516	85.7%
Noninterviews	7,794	2,909,466	2.5%
Vacants	28,472	10,398,118	9.0%
Deletes	10,472	3,176,108	2.7%
Interview rate[2]	0.97	0.97	N/A

[1] The A.C.E. household interview is an interview (for Census Day or A.C.E. Interview Day) if there is at least one person (with name and at least two demographic characteristics) who possibly or definitely was a resident of the housing unit on Census Day or A.C.E. Interview Day.

[2] The unweighted (weighted) interview rate is the unweighted (weighted) number of interviews divided by the unweighted (weighted) sum of interviews and noninterviews.

Census Day Final Interview Outcome

The final census day outcome codes are in Table A-27. Changes as a result of the follow-up interview are

- Whole households of P-sample people who said they lived elsewhere on census day are converted to a noninterviews.

- Whole households who lived in group quarters on census day or should have been enumerated at another residence are converted to vacant.

Table A-27. Census Day Final Estimation Outcome Codes for P-sample Housing Units (Weighted)

Outcome code	Housing units (weighted)	Percent (weighted)
Interviews	**99,166,516**	**85.7%**
Complete interview with a household member (1)	91,323,049	79.0%
Complete interview with a proxy respondent (2)	6,876,428	5.9%
Partial interview (3)	967,039	0.8%
Noninterviews	**2,909,466**	**2.5%**
No census day residents - household converted to noninterview (4)	948,550	0.8%
Field noninterview (6)	1,002,590	0.9%
All people have insufficient information for matching and follow-up (9)	958,326	0.8%
Vacants	**10,398,118**	**9.0%**
No census day residents - vacant (10)	1,700,431	1.5%
Vacant on census day (11)	8,697,687	7.5%
Deletes - Not a housing unit on census day (12)	**3,176,108**	**2.7%**
Total	**115,650,208**	**100.0%**

Table 1b. Status of A.C.E. Household Interviews[3] for A.C.E. Interview Day: 2000 Data

	Number (unweighted)	Number (weighted)	Percent (weighted)
U.S. Total (Housing Units)	300,913	115,650,208	100.0%
Interviews	264,103	102,651,540	88.8%
Noninterviews	3,052	1,196,445	1.0%
Vacants	29,662	10,527,420	9.1%
Deletes	4,096	1,274,803	1.1%
Interview rate[4]	0.99	0.99	N/A

Table 1c. Status of PES Household Interviews[3] (for the Day of the PES Interview): 1990 Data **

1990 PES	Number (unweighted)		Number (weighted)	Percent (weighted)
U.S. Total (Housing Units)	171,390			
Interviews	141,667	(82.66%)		
Noninterviews	2,246	(1.31%)		
Vacants & Deletes	27,477	(16.03%)		
Interview rate[4]		98.4%		

** Checking to see if data are available to create this table.

[3] The A.C.E. household interview is an interview (for Census Day or A.C.E. Interview Day) if there is at least one person (with name and at least two demographic characteristics) who possibly or definitely was a resident of the housing unit on Census Day or A.C.E. Interview Day.

[4] The unweighted (weighted) interview rate is the unweighted (weighted) number of interviews divided by the unweighted (weighted) sum of interviews and noninterviews.

A.C.E. Interview Day Final Interview Outcome

The final interview day outcome codes are in Table A-28. The interview outcome as of A.C.E. interview day is for the non-movers and the in-movers. Changes as a result of the follow-up interview are in whole households of non-movers who said they lived elsewhere, in group quarters, or have another residence where they should have been counted on census day are converted to noninterviews.

Table A-28. A.C.E. Interview Day Final Estimation Outcome Codes for P-sample Housing Units (Weighted)

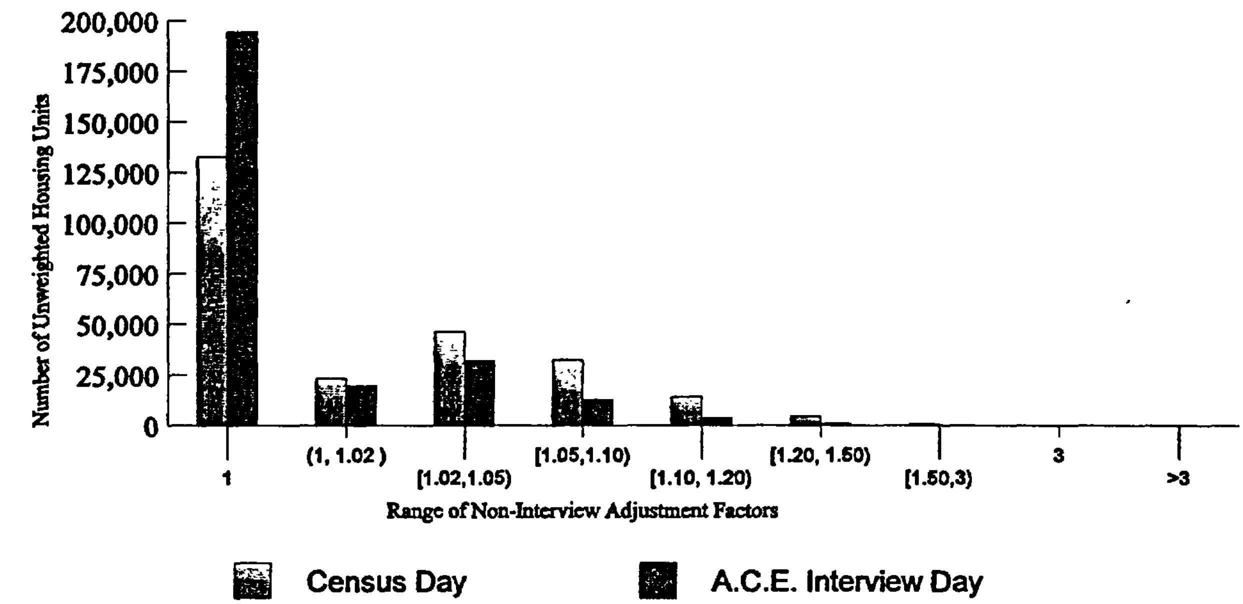
Outcome code	Housing units (weighted)	Percent (weighted)
Interviews	**102,651,540**	**88.8%**
Complete interview with a household member (1)	97,280,209	84.1%
Complete interview with a proxy respondent (2)	4,624,302	4.0%
Partial interview (3)	747,029	0.6%
Noninterviews	**1,196,445**	**1.0%**
Field noninterview	135,272	
Refusal (5)	66,066	0.1%
Unable to contact knowledgeable respondent (7)	68,933	0.1%
Language problems (8)	273	0.0%
All people have insufficient information for matching and follow-up (9)	880,308	0.8%
No A.C.E. interview day residents - household converted to noninterview (10)	180,865	14.2%
Vacants - Vacant on A.C.E. interview day (11)	**10,527,420**	**9.1%**
Deletes - Not a housing unit on A.C.E. interview day (12)	**1,274,803**	**1.1%**
Total	**115,650,208**	**100.0%**

Table 2 shows the distribution of noninterview adjustment factors for census day and A.C.E. interview day. This table allows us to quickly observe one facet of nonresponse bias on A.C.E. data, because noninterview adjustment factors are a function of the nonresponse rate. The histogram below Table 2 summarizes the same information in graphical format.

Table 2. Distribution of Noninterview Adjustment Factors for Census Day and A.C.E. Interview Day

	Number of housing units with corresponding noninterview adjustment factor (Unweighted)									
	1	(1, 1.02)	[1.02, 1.05)	[1.05, 1.10)	[1.10, 1.20)	[1.20, 1.50)	[1.50,3)	3	>3	Total HU
Census day	132,828	23,173	46,328	32,382	14,089	4,582	791	2	0	254,175
Percent	52.3%	9.1%	18.2%	12.7%	5.5%	1.8%	0.3%	0.0%	0.0%	100.0%
A.C.E. interview day	194,430	19,776	32,179	12,965	3,713	908	132	0	0	264,103
Percent	73.6%	7.5%	12.2%	4.9%	1.4%	0.3%	0.0%	0.0%	0.0%	100.0%

Table 2a. Distribution of Census Day and A.C.E. Day Non-Interview Adjustment Factors



Characteristic Imputation

At times, people in the P and E Samples are missing one or more of the following characteristics on the census or A.C.E. questionnaires: age, sex, tenure, race, or Hispanic origin. When missing, each of these items must be imputed so that the person can be assigned to a post-stratum for dual system estimation. Characteristic imputation is not carried out for other missing variables (with the exception of the unresolved status items discussed later) as they are not needed to determine post-stratification. The imputation methods for the P Sample and the E Sample differ, as each has different sources of data available to use for imputation.

P-Sample characteristic imputation for Census 2000 is nearly identical to that for the Dress Rehearsal. Imputation for a specific missing characteristic in the P Sample is not affected by the imputation for other missing characteristics. That is, the algorithms are applied independently. Before imputation begins, age and sex distributions are calculated nationally using the P-Sample data. Missing age or sex is then drawn from the appropriate conditional distribution. Tenure, race, and Hispanic origin are imputed essentially using a hot-deck procedure, where the data are sorted by cluster, then map spot number, then unit identifier. This essentially produces a geographic sort of the data file. Mover status for P-Sample people is not considered when imputing characteristics. Details are found in DSSD Census 2000 Procedures and Operations Memorandum Series #Q-19 and #Q-25.

For a missing characteristic in the Census 2000 E Sample, whenever possible we use the actual value of the characteristic imputed in the census. That is, we match the E-Sample person record to its counterpart on the edited file for the entire 2000 Census, and extract the missing characteristic. In the unlikely event that an E-Sample record is still missing a required characteristic, the corresponding P-Sample procedure would be used.

NOTE: In the 1990 PES procedures for characteristic imputation for the E Sample, we actually imputed values for missing data separately from the census imputation; therefore, the E-Sample rates of imputation for the 2000 A.C.E. and for the1990 PES are not directly comparable.

Table 3b. A.C.E. Characteristic Imputation Percent Rates for the P Sample by Proxy and Mover Status and for the E Sample: 2000 Data (Weighted frequencies and rates)

	Total people	Percentage of people with imputed characteristic					Percentage of people with 1 or more imputed characteristics
		Age	Sex	Tenure	Race	Hispanic origin	
P sample							
Total	281,708,154	2.4%	1.7%	1.9%	1.4%	2.3%	5.4%
Proxy status							
Non-proxy	265,781,888	2.1%	1.5%	1.7%	1.0%	1.8%	4.4%
Proxy	15,926,266	7.9%	4.2%	5.2%	8.7%	11.0%	21.9%
Mover status							
Non-mover	258,455,070	2.3%	1.7%	1.9%	1.2%	2.1%	5.0%
In-mover	13,571,043	2.3%	0.4%	0.4%	1.3%	0.8%	3.7%
Out-mover	9,682,040	6.0%	3.4%	2.3%	8.0%	9.0%	17.4%
E sample							
Total	264,578,862	2.9%	0.2%	3.6%	3.2%	3.4%	13.0%

Table 3d. PES Characteristic Imputation Percent Rates for the P Sample by Proxy and Mover Status and for the E Sample: 1990 Data (Weighted frequencies and rates)

1990 PES	Total people	Percentage of people with imputed characteristic					Percentage of people with 1 or more imputed characteristics
		Age	Sex	Tenure	Race	Hispanic origin	
P sample							
Total	240,651,222	0.71%	0.51%	2.26%	2.49%		
Proxy status							
Non-proxy							
Proxy							
Mover status[9]							
Non-mover							
In-mover							
E sample[10] **							
Total	244,200,930	2.39%	1.04%	2.48%	11.75%		

** Checking to see if data are available to create this table.

Data from 1990 PES Evaluation Project P2: Distribution of Missing Data Rates, Table 3.3, p. 11.

[9] Out-movers are not included in these tables for the 1990 PES because Procedure B was used.

[10] In the 1990 PES, characteristic imputation for the E Sample was done separately from the census imputation; therefore, the E-Sample imputation rates for 2000 A.C.E. and for the 1990 PES are not directly comparable.

Tables 4a and 4b show the missing data rates for each post-stratification variable, along with the distribution of the item responses before and after imputation for the weighted P sample, weighted E sample, and the entire census.

Table 4a. Distribution of Characteristics Before and After Item Imputation (Race and Hispanic Origin)

Response	P sample (weighted)		E sample (weighted)		Census *	
	Before imputation	After imputation	Before imputation	After imputation	Before imputation	After imputation
Total	281,708,154	281,708,154	264,578,862	264,578,862	ERR	ERR
Race **	Missing Race 1.4%		Missing Race 3.2%		ERR	ERR
White only	73.5%	73.4%	76.9%	76.2%	ERR	ERR
Black only	11.0%	11.0%	11.8%	11.6%	ERR	ERR
AIAN[13] only	0.6%	0.6%	0.8%	0.8%	ERR	ERR
Asian only	3.5%	3.5%	3.7%	3.7%	ERR	ERR
NHPI[14] only	0.1%	0.1%	0.1%	0.1%	ERR	ERR
Other race only	8.3%	8.4%	4.5%	5.3%	ERR	ERR
Multi-race	3.0%	3.0%	2.3%	2.3%	ERR	ERR
Hispanic origin **	Missing Hisp. Origin 2.3%		Missing Hisp. Origin 3.4%		ERR	ERR
Hispanic	12.4%	12.4%	12.5%	12.4%	ERR	ERR
Non-Hispanic	87.6%	87.6%	87.5%	87.6%	ERR	ERR

* Checking to see if data are available to create this table.

** The weighted percent missing for each characteristic is the weighted number of people with item missing data for that characteristic divided by the total weighted number of people.

The weighted percent for each category before characteristic imputation is the weighted number of people in the category divided by the total weighted number of people excluding the weighted number of people with item missing data for that characteristic.

13 American Indian or Alaskan Native

14 Native Hawaiian or Pacific Islander

Table 4b. Distribution of Characteristics Before and After Item Imputation (Age, Sex, and Tenure)

Response	P sample (weighted)		E sample (weighted)		Census *	
	Before imputation	After imputation	Before imputation	After imputation	Before imputation	After imputation
Total	281,708,154	281,708,154	264,578,862	264,578,862	ERR	ERR
Age **	Missing Age 2.4%		Missing Age 2.9%		ERR	ERR
0 - 17	26.1%	26.0%	25.9%	25.7%	ERR	ERR
18 - 29	16.7%	16.7%	15.5%	15.6%	ERR	ERR
30 - 49	30.7%	30.8%	31.0%	31.0%	ERR	ERR
50 +	26.5%	26.5%	27.6%	27.6%	ERR	ERR
Sex **	Missing Sex 1.7%		Missing Sex 0.2%		ERR	ERR
Male	48.4%	48.3%	48.8%	48.8%	ERR	ERR
Female	51.6%	51.7%	51.2%	51.2%	ERR	ERR
Tenure **	Missing Tenure 1.9%		Missing Tenure 3.6%		ERR	ERR
Owner	68.4%	68.4%	69.9%	69.7%	ERR	ERR
Non-owner	31.6%	31.6%	30.1%	30.3%	ERR	ERR

* Checking to see if data are available to create this table.

** The weighted percent missing for each characteristic is the weighted number of people with item missing data for that characteristic divided by the total weighted number of people.

The weighted percent for each category before characteristic imputation is the weighted number of people in the category divided by the total weighted number of people excluding the weighted number of people with item missing data for that characteristic.

Imputation of Status (Resident, Match, Correct Enumeration)

After all follow-up activities are completed, there remains a small fraction of the A.C.E. sample people for whom we still do not have enough information to compute the components of the dual system estimator. Their status is said to be "unresolved." We use imputation cell estimation to assign probabilities for P-Sample people with unresolved match or Census-Day resident status, and for E-Sample people with unresolved enumeration status.

All P- and E-Sample people–resolved and unresolved–are separated into groups called imputation cells based on operational and demographic characteristics. We use different variables to define cells for P- and E-Sample people, and, among P-Sample people, to define cells for resolving match and resident status. Within each imputation cell the weighted proportion of matches (or residents or correct enumerations) among the cases with resolved status is calculated, and that value is imputed for all unresolved people in the cell.
Note: Some people are removed from the P Sample, including people who are considered to be fictitious, duplicates, geocoding errors, or not residents of the housing unit on census day. These people are not included in the following tables related to resident or match status.

Results - Imputation of Status

Resident Status

Tables 5a and 5b respectively show the unweighted and weighted final resident status for P-Sample people for the U.S. in the 2000 A.C.E. by mover status and by region. These tables also show the resident rates for resolved cases. Note that no people actually received this rate because this rate is a weighted average over the all imputation classes. The final P-Sample resident status is broken down into

- (confirmed) resident
- (confirmed) nonresident
- unresolved resident status

Resident - The matched or nonmatched P-sample person is a resident of the housing unit on Census Day.

Nonresident - P-sample people are nonresidents of the cluster when they are fictitious, duplicates, geocoding errors, or should not have been included as a resident of the housing unit on census day. Nonresidents are removed from the P sample.

Unresolved Resident Status - A matched or nonmatched P-sample person has unresolved resident status when the follow-up interview did not successfully determine the person's residence on census day. The resident status of the possible match is unresolved when the follow-up interview was not successful. The resident status is also imputed when the P-sample person had insufficient information for matching.

For Tables 5a, 5b, 8, A-9, and A-15 through A-18, the weighted resident rate is determined by dividing the weighted number of confirmed residents by the weighted number of resolved cases--all confirmed residents and nonresidents. When calculating this rate, we only include people with mover status of non-mover and out-mover.

(By definition, non-movers and out-movers should both be Census Day residents; however, we create the mover-status variable prior to field follow-up work. This work may reveal that a non-mover or out-mover was not actually a Census Day resident. For example, a person may report he or she lived in the housing unit since March 20. Preliminary operations would label this person a non-mover; however, follow-up operations may confirm this person moved into the housing unit on April 20. Therefore, this person is a confirmed non-resident for Census day.)

Table 5a. Final Resident Status for the P Sample in the A.C.E. by Mover Status and by Region
(Unweighted frequencies and rates)

P sample	Total people	Final resident status: Confirmed resident		Confirmed nonresident		Unresolved resident status		Resident rate for resolved cases
U.S. Total[15]	653,338	625,863	95.8%	12,393	1.9%	15,082	2.3%	0.98
Mover status								
Non-mover	627,992	606,816	96.6%	10,502	1.7%	10,674	1.7%	0.98
In-mover	1	0	0.0%	1	100.0%	0	0.0%	0.00
Out-mover	25,345	19,047	75.2%	1,890	7.5%	4,408	17.4%	0.91
Region								
Northeast	122,226	117,369	96.0%	2,566	2.1%	2,291	1.9%	0.98
Midwest	145,410	140,819	96.8%	2,053	1.4%	2,538	1.7%	0.99
South	209,197	199,111	95.2%	4,502	2.2%	5,584	2.7%	0.98
West	176,505	168,564	95.5%	3,272	1.9%	4,669	2.6%	0.98

[15] This total excludes 15,489 P-Sample TES people with TES weight = 0; they were not selected for the TES sample.

Table 5b. Final Resident Status for the P Sample in the A.C.E. by Mover Status and by Region
(Weighted frequencies and rates)

P sample	Total people	Final resident status						Resident rate for resolved cases
		Confirmed resident		Confirmed nonresident		Unresolved resident status		
U.S. Total	257,479,497	247,353,379	96.1%	4,452,044	1.7%	5,674,074	2.2%	0.98
Mover status								
Non-mover	247,868,634	240,077,899	96.9%	3,773,365	1.5%	4,017,370	1.6%	0.98
In-mover	275	0	0.0%	275	100.0%	0	0.0%	
Out-mover	9,610,587	7,275,480	75.7%	678,404	7.1%	1,656,704	17.2%	0.91
Region								
Northeast	49,064,972	47,260,156	96.3%	949,002	1.9%	855,814	1.7%	0.98
Midwest	58,913,626	57,170,167	97.0%	774,946	1.3%	968,514	1.6%	0.99
South	90,777,640	86,658,784	95.5%	1,845,380	2.0%	2,273,476	2.5%	0.98
West	58,723,259	56,264,272	95.8%	882,716	1.5%	1,576,270	2.7%	0.98

See Table A-9 in the Appendix for weighted frequencies and rates for P-sample final resident status by state.

Match Status

As with resident status, the match status of a P-sample person may be unresolved despite the attempts of field follow-up. Tables 6a and 6b, respectively, show the unweighted and weighted final match status for P-Sample people for the U.S. in the 2000 A.C.E. by mover status and by region. These tables also show match rates for resolved cases. The final P-Sample match status is broken down into

- match,
- nonmatch,
- unresolved match status

Match - The P-sample person was found in the E-Sample listing of the search area; that is, in the cluster or in the surrounding block in either a housing unit or in group quarters.

Nonmatch - The P-sample person was not found in the search area. If the nonmatch was sent to follow-up, the person was confirmed to be a resident of the cluster on census day. If the nonmatch was not sent for a follow-up interview, a household member identified the person as a resident of the housing unit during the original ACE interview.

Unresolved match status - The match status is unresolved for possible matches with unsuccessful follow-up interviews and for P-sample people with insufficient information for matching and follow-up.

For Tables 6a through 6d, 9, A-10, A-19 through A-22, and A-29 through A-39, the weighted match rate is determined by dividing the weighted number of matches by the weighted number of resolved cases--the sum of matches and nonmatches. To calculate the probability of match status, we only consider Census Day confirmed residents and people with unresolved resident status. That is, we exclude confirmed non-residents while calculating match probabilities.

For comparison, Tables 6c and 6d contain data from the 1990 PES.

Table 6a. Final Match Status for the P Sample in the A.C.E. by Mover Status and by Region: 2000 Data
(Unweighted frequencies and rates)

P sample	Total people	Final match status						Match rate for resolved cases
		Match		Nonmatch		Unresolved match		
U.S. Total[16]	640,945	578,695	90.3%	54,424	8.5%	7,826	1.2%	0.91
Mover status								
Non-mover	617,490	562,783	91.1%	49,345	8.0%	5,362	0.9%	0.92
Out-mover	23,455	15,912	67.8%	5,079	21.7%	2,464	10.5%	0.76
Region								
Northeast	119,660	107,832	90.1%	10,450	8.7%	1,378	1.2%	0.91
Midwest	143,357	133,073	92.8%	8,937	6.2%	1,347	0.9%	0.94
South	204,695	182,833	89.3%	19,151	9.4%	2,711	1.3%	0.91
West	173,233	154,957	89.5%	15,886	9.2%	2,390	1.4%	0.91

Table 6b. Final Match Status for the P Sample in the A.C.E. by Mover Status and by Region: 2000 Data
(Weighted frequencies and rates)

P sample	Total people	Final match status						Match rate for Resolved Cases
		Match		Nonmatch		Unresolved match		
U.S. total	253,027,452	229,196,468	90.6%	20,780,055	8.2%	3,050,930	1.2%	0.92
Mover status								
Non-mover	244,095,269	223,123,506	91.4%	18,844,222	7.7%	2,127,541	0.9%	0.92
Out-mover	8,932,183	6,072,962	68.0%	1,935,832	21.7%	923,389	10.3%	0.76
Region								
Northeast	48,115,970	43,605,704	90.6%	3,988,908	8.3%	521,358	1.1%	0.92
Midwest	58,138,680	54,069,501	93.0%	3,543,936	6.1%	525,243	0.9%	0.94
South	88,932,259	79,596,127	89.5%	8,183,539	9.2%	1,152,593	1.3%	0.91
West	57,840,542	51,925,135	89.8%	5,063,671	8.8%	851,737	1.5%	0.91

[16] This total excludes 15,127 P-Sample TES people with TES weight = 0; they were not selected for the TES sample.

Enumeration Status

Tables 7a and 7b, respectively, show the unweighted and weighted final enumeration status for the E Sample in the 2000 A.C.E. for the total U.S. and by region. These tables consist of the Total People (number of people in the E Sample), the Final Enumeration Status, and the Correct Enumeration Rate for Resolved Cases. The final E-Sample enumeration status is broken down

- correct enumeration
- erroneous enumeration
- unresolved enumeration status

Correct Enumeration - E-sample people are correctly enumerated when they are matched to the P-sample or when they have been followed up and they should have been enumerated in this cluster.

Erroneous Enumeration - E-sample people are erroneously enumerated when they have another residence where they should be counted on census day, are fictitious, are duplicated, lived in a housing unit that was a geocoding error, or have insufficient information for matching and follow-up.

Unresolved Enumeration Status - E-sample people have unresolved enumeration status when the follow-up interview was unsuccessful. The E-sample person may have been followed up to obtain information about the E-sample nonmatch, possible match, matched person with unresolved resident status, or geographic work to obtain the location of the housing unit.

For Tables 7a through 7d, 10, and A-11, A-23 through A-26, and A-29 through A-39, the weighted enumeration rate is determined by dividing the weighted number of correctly enumerated people by the weighted number of resolved cases (the sum of correctly enumerated and erroneously enumerated people).

For comparison, Tables 7c and 7d contain data from the 1990 PES.

Table 7a. Final Enumeration Status for the E Sample in the A.C.E. by Region: 2000 Data
(Unweighted frequencies and rates)

E sample	Total people	Final enumeration status						Correct enumeration rate for resolved cases
		Correct enumeration		Erroneous enumeration		Unresolved enumeration		
U.S. Total[19]	704,602	652,390	92.6%	31,064	4.4%	21,148	3.0%	0.95
Region								
Northeast	1[illegible]2,67[illegible]	122,968	92.7%	6,841	5.2%	2,864	2.2%	0.95
Midwest	152,315	143,660	94.3%	5,303	3.5%	3,352	2.2%	0.96
South	231,503	212,514	91.8%	10,720	4.6%	8,269	3.6%	0.95
West	188,111	173,248	92.1%	8,200	4.4%	6,663	3.5%	0.95

Table 7b. Final Enumeration Status for the E Sample in the A.C.E. by Region: 2000 Data
(Weighted frequencies and rates)

E sample	Total people	Final enumeration status						Correct enumeration rate for resolved cases
		Correct enumeration		Erroneous enumeration		Unresolved enumeration		
U.S. Total	264,578,862	246,999,032	93.4%	10,688,934	4.0%	6,890,897	2.6%	0.96
Region								
Northeast	50,436,513	47,192,325	93.6%	2,293,348	4.5%	950,840	1.9%	0.95
Midwest	60,196,168	56,985,634	94.7%	1,954,872	3.2%	1,255,661	2.1%	0.97
South	94,296,537	87,529,893	92.8%	4,034,026	4.3%	2,732,618	2.9%	0.96
West	59,649,645	55,291,181	92.7%	2,406,687	4.0%	1,951,778	3.3%	0.96

19 This total excludes 8,298 E-Sample TES people with TES weight = 0; they were not selected for the TES sample.

Table 8. Imputation Cells Used for P-Sample Resident Status (Weighted Frequencies and Rates)

Before follow-up match code group	Final resident status (weighted)	Owner: Non-Hispanic White only	Owner: Others	Non-owner: Non-Hispanic White only	Non-owner: Others	Total
1 = Matches needing follow-up	Total people	1,066,783	275,841	474,578	327,786	2,144,988
	Confirmed residents	954,103	255,772	351,327	249,731	1,810,933
	Confirmed nonresidents	17,834	3,716	2,405	2,193	26,148
	Unresolved resident status	94,846	16,353	120,845	75,863	307,907
	Resident rate for resolved cases	0.98	0.99	0.99	0.99	0.99
2 = Possible matches	Total people	548,089	406,432	323,301	579,803	1,857,625
	Confirmed residents	503,726	374,013	290,322	513,214	1,681,275
	Confirmed nonresidents	14,092	12,234	10,303	14,520	51,149
	Unresolved resident status	30,271	20,184	22,676	52,070	125,201
	Resident rate for resolved cases	0.97	0.97	0.97	0.97	0.97
3 = Partial household nonmatches needing follow-up V3a - Age 18-29 and Child of reference person	Total people	268,508	144,560	54,002	106,711	573,781
	Confirmed residents	191,115	125,796	42,168	89,720	448,798
	Confirmed nonresidents	61,862	13,824	5,588	6,977	88,250
	Unresolved resident status	15,531	4,940	6,247	10,015	36,733
	Resident rate for resolved cases	0.76	0.90	0.88	0.93	0.84
3 = Partial household nonmatches needing follow-up V3b - Others	Total people	1,608,515	1,229,905	761,823	1,532,235	5,132,478
	Confirmed residents	1,433,998	1,133,145	614,907	1,345,939	4,527,989
	Confirmed nonresidents	66,420	34,181	26,222	42,476	169,299
	Unresolved resident status	108,097	62,579	120,693	143,821	435,191
	Resident rate for resolved cases	0.96	0.97	0.96	0.97	0.96
4 = Whole household nonmatches needing follow-up, not conflicting households	Total people	1,193,507	319,154	866,955	754,555	3,134,170
	Confirmed residents	995,407	277,479	637,280	562,795	2,472,962
	Confirmed nonresidents	86,886	16,648	62,468	52,734	218,736
	Unresolved resident status	111,213	25,027	167,207	139,026	442,472
	Resident rate for resolved cases	0.92	0.94	0.91	0.91	0.92
5 = Nonmatches from conflicting households	Total people	557,394	426,811	578,019	919,900	2,482,124
	Confirmed residents	393,842	331,395	429,145	698,409	1,852,791
	Confirmed nonresidents	38,948	26,150	24,825	33,383	123,306
	Unresolved resident status	124,605	69,266	124,049	188,108	506,028
	Resident rate for resolved cases	0.91	0.93	0.95	0.95	0.94
6 = Resolved before follow-up	Total people	131,529,056	37,306,611	35,224,222	32,301,327	236,361,216
	Confirmed residents	130,608,330	36,923,966	34,854,927	31,889,265	234,276,488
	Confirmed nonresidents	910,784	375,966	359,223	391,344	2,037,317
	Unresolved resident status	9,942	6,680	10,072	20,717	47,411
	Resident rate for resolved cases	0.99	0.99	0.99	0.99	0.99
7 = Insufficient information for matching	Total people	1,162,840	460,854	686,196	625,186	2,935,076
	Confirmed residents	1,074	491	0	0	1,565
	Confirmed nonresidents	200	0	2,519	0	2,719
	Unresolved resident status	1,161,566	460,363	683,677	625,186	2,930,792
	Resident rate for resolved cases	0.81	0.87	0.84	0.87	0.84
8 = Potentially fictitious or people said to be living elsewhere on census day	Total people	1,096,721	436,653	591,119	733,545	2,858,038
	Confirmed residents	104,111	39,750	68,233	68,485	280,579
	Confirmed nonresidents	767,180	283,468	317,283	367,190	1,735,121
	Unresolved resident status	225,431	113,435	205,603	297,870	842,339
	Resident rate for resolved cases	0.12	0.12	0.18	0.16	0.14
Total	Total people	139,031,413	41,006,821	39,560,215	37,881,048	257,479,497
	Confirmed residents	135,185,706	39,461,807	37,288,309	35,417,557	247,353,379
	Confirmed nonresidents	1,964,205	766,187	810,837	910,816	4,452,044
	Unresolved resident status	1,881,502	778,826	1,461,070	1,552,675	5,674,074
	Resident rate for resolved cases	0.99	0.98	0.98	0.97	0.98

Table 9. Imputation Cells Used for P-Sample Match Status (Weighted Frequencies and Rates)

Mover status	Final match status (Weighted)	Housing-unit address match code: Housing unit was a match (code 1)[20]		Housing-unit address match code: Housing unit was a nonmatch or the household is conflicting (code 2 or 4)		Total
		No imputes	1 or more imputes	No imputes	1 or more imputes	
Non-mover	Total people	212,844,647	11,078,431	19,061,019	1,111,173	244,095,269
	Matched	200,421,491	9,092,884	13,060,089	549,042	223,123,506
	Nonmatched	11,562,492	998,717	5,863,101	419,912	18,844,222
	Unresolved match	860,663	986,831	137,829	142,219	2,127,541
	Match rate for resolved cases	0.95	0.90	0.69	0.57	0.92
Out-mover	Total people	6,416,631	1,263,115	1,252,438		8,932,183
	Matched	4,746,267	756,997	569,698		6,072,962
	Nonmatched	1,200,572	200,506	534,755		1,935,832
	Unresolved match	469,792	305,612	147,985		923,389
	Match rate for resolved cases	0.80	0.79	0.52		0.76
Total	Total people	219,261,277	12,341,546	21,424,629		253,027,452
	Matched	205,167,758	9,849,881	14,178,829		229,196,468
	Nonmatched	12,763,064	1,199,222	6,817,768		20,780,055
	Unresolved match	1,330,455	1,292,443	428,032		3,050,930
	Match rate for resolved cases	0.94	0.89	0.68		0.92

20 The P-Sample Address Codes used for creating these cells are (from DSSD Memo #Q-25)

1 = Housing Unit Matched during Housing Unit Matching
2 = Housing Unit Did Not Match during Housing Unit Matching
4 = Conflicting Households

Address code values of 2-3 are considered to be "Housing Units not matched during Housing Unit matching" for the purposes of match code group assignment.

Table 10. Imputation Cells Used for E-Sample Enumeration Status (Weighted Frequencies and Rates)

Before follow-up group	Final enumeration status	No imputes		1 or more imputes	Total
1 = Matches needing follow-up	Total people		1,737,638	302,752	2,040,390
	Correct enumeration		1,467,910	237,683	1,705,593
	Erroneous enumeration		35,121	5,637	40,758
	Unresolved enumeration		234,607	59,432	294,039
	Correct enum. rate for resolved cases		0.98	0 98	0.98
2 = Possible matches	Total people		1,204,233	505,884	1,710,117
	Correct enumeration		1,092,888	457,838	1,550,726
	Erroneous enumeration		36,263	14,898	51,161
	Unresolved enumeration		75,083	33,147	108,230
	Correct enum. rate for resolved cases		0.97	0.97	0.97
3 = Partial household nonmatches V3a - Age 18-29 and Child of Reference Person	Total people		898,638	145,705	1,044,343
	Correct enumeration		725,012	117,959	842,970
	Erroneous enumeration		107,176	11,894	119,071
	Unresolved enumeration		66,450	15,852	82,302
	Correct enum. rate for resolved cases		0.87	0.91	0.88
3 = Partial household nonmatches V3b - Others	Total people		5,954,242	1,336,054	7,290,295
	Correct enumeration		5,248,294	1,149,350	6,397,644
	Erroneous enumeration		138,322	48,098	186,420
	Unresolved enumeration		567,626	138,605	706,231
	Correct enum. rate for resolved cases		0.97	0.96	0.97
		Non-Hispanic White	Others		
4 = Whole household nonmatches where the housing unit matched; not conflicting households	Total people	5,174,133	2,816,066	1,563,432	9,553,631
	Correct enumeration	4,230,898	2,229,265	1,256,996	7,717,158
	Erroneous enumeration	152,741	60,466	55,462	268,669
	Unresolved enumeration	790,493	526,335	250,975	1,567,804
	Correct enum. rate for resolved cases	0.97	0.97	0.96	0.97
5 = Nonmatches from conflicting households; for housing units not in regular nonresponse follow-up	Total people		912,138	200,054	1,112,191
	Correct enumeration		730,314	155,315	885,629
	Erroneous enumeration		18,405	5,631	24,036
	Unresolved enumeration		163,418	39,108	202,526
	Correct enum. rate for resolved cases		0.98	0.97	0.97
6 = Nonmatches from conflicting households; housing units in regular nonresponse follow-up	Total people		841,318	259,888	1,101,206
	Correct enumeration		585,138	175,327	760,465
	Erroneous enumeration		55,275	14,014	69,289
	Unresolved enumeration		200,906	70,547	271,453
	Correct enum. rate for resolved cases		0.91	0.93	0.92
		Non-Hispanic White	Others		
7 = Whole household nonmatches, where the housing unit did not match in housing unit matching	Total people	4,932,503	2,042,887	1,129,229	8,104,620
	Correct enumeration	4,332,845	1,713,340	952,272	6,998,457
	Erroneous enumeration	184,679	96,780	50,230	331,689
	Unresolved enumeration	414,979	232,767	126,728	774,474
	Correct enum. rate for resolved cases	0.96	0.95	0.95	0.95

Table 10. (Cont.) Imputation Cells Used for E-Sample Enumeration Status (Weighted Frequencies and Rates)

Before follow-up group	Final enumeration status	No imputes		1 or more imputes	Total
		Non-Hispanic White	Others		
8 = Resolved before follow-up	Total people	142,798,309	47,919,570	23,769,194	214,487,072
	Correct enumeration	142,078,256	47,397,936	23,262,689	212,738,881
	Erroneous enumeration	692,617	502,747	492,335	1,687,698
	Unresolved enumeration	27,435	18,887	14,170	60,492
	Correct enum. rate for resolved cases	1.00	0.99	0.98	0.99
9 = Insufficient information for matching	Total people		1,442,884	3,317,320	4,760,204
	Correct enumeration		4,542	4,029	8,571
	Erroneous enumeration		1,436,980	3,312,434	4,749,414
	Unresolved enumeration		1,362	857	2,219
	Correct enum. rate for resolved cases		0.00	0.00	0.00
10 = TES people	Total people		7,533,567	957,446	8,491,013
	Correct enumeration		6,047,941	703,156	6,751,097
	Erroneous enumeration		472,275	116,090	588,365
	Unresolved enumeration		1,013,351	138,200	1,151,551
	Correct enum. rate for resolved cases		0.93	0.86	0.92
11 = Potentially fictitious people	Total people		871,606	215,606	1,087,212
	Correct enumeration		23,276	9,534	32,810
	Erroneous enumeration		377,991	98,501	476,492
	Unresolved enumeration		470,338	107,572	577,910
	Correct enum. rate for resolved cases		0.06	0.09	0.06
12 = People who were said to be living elsewhere on census day	Total people		3,070,983	725,585	3,796,568
	Correct enumeration		501,046	107,984	609,031
	Erroneous enumeration		1,689,200	406,671	2,095,872
	Unresolved enumeration		880,737	210,929	1,091,666
	Correct enum. rate for resolved cases		0.23	0.21	0.23
TOTAL	Total people		230,150,714	34,428,149	264,578,862
	Correct enumeration		218,408,901	28,590,131	246,999,032
	Erroneous enumeration		6,057,038	4,631,896	10,688,934
	Unresolved enumeration		5,684,775	1,206,122	6,890,897
	Correct enum. rate for resolved cases		0.97	0.86	0.96

Table 3. P-Sample Person Counts for Missing Data (Unweighted)

1/08/01

	Total	Non-TES People	TES People (TESWGT > 0)	TES People (TESWGT = 0)	Overall Total
Total	706,245	673,192	33,053	15,489	721,734
Resident Set	653,338	620,285	33,053	15,489	668,827
Match Set (Confirmed or Possible Residents)	640,945	608,523	32,422	15,127	656,072
Matches	578,695	556,815	21,880	0	578,695
Confirmed Residents					
M	567,091	545,927	21,164	0	567,091
MR	10,553	9,934	619	0	10,553
Unresolved Residents					
MU	1,051	954	97	0	1,051
Nonmatches	54,424	43,890	10,534	15,127	69,551
Confirmed Residents					
NR	27,042	22,875	4,167	4,794	31,836
NP	19,746	14,001	5,745	9,884	29,630
NC	1,431	1,431	0	0	1,431
Unresolved Residents					
NU	6,205	5,583	622	449	6,654
Unresolved Matches (All Unresolved Residents)	7,826	7,818	8	0	7,826
KI	3,994	3,994	0	0	3,994
KP	3,699	3,699	0	0	3,699
P	133	125	8	0	133
Confirmed Nonresidents	12,393	11,762	631	362	12,755
Matches					
MN	205	198	7	0	205
Nonmatches					
FP	603	577	26	35	638
NL	2,529	2,366	163	68	2,597
NN	2,529	2,327	202	101	2,630
Other					
DP	6,175	6,175	0	0	6,175
GP	352	119	233	158	510
Not in Resident Set (No Match Codes)	52,907	52,907	0	0	52,907
RSC = I	36,656	36,656	0	0	36,656
MOVERPER = 2	36,622	36,622	0	0	36,622
MOVERPER = 1	34	34	0	0	34
RSC = R	16,203	16,203	0	0	16,203
RSC = N or U (POC = 9; FOC = 9)	48	48	0	0	48

Table 5. P-Sample Person Counts for Estimation (Unweighted)

1/08/01

		Total	Non-TES People	TES People (TESWGT > 0)	TES People (TESWGT = 0)	Overall Total
Total		**706,245**	**673,192**	**33,053**	**15,489**	**721,734**
Non-movers (MOVERPER=1, RSC≠R)		**628,074**	**596,932**	**31,142**	**14,741**	**642,815**
Census Day Resident Prob. = 1		**606,816**	**576,590**	**30,226**	**14,090**	**620,906**
Matches						
M		553,152	532,799	20,353	0	553,152
MR		9,114	8,579	535	0	9,114
Nonmatches						
NR		24,416	20,621	3,795	4,457	28,873
NP		18,782	13,239	5,543	9,633	28,415
NC		1,352	1,352	0	0	1,352
Census Day Resident Prob. (0,1)		**10,674**	**10,131**	**543**	**334**	**11,008**
Matches	MU	517	466	51	0	517
Nonmatches	NU	4,795	4,311	484	334	5,129
Other						
KI		2,595	2,595	0	0	2,595
KP		2,653	2,653	0	0	2,653
P		114	106	8	0	114
Census Day Resident Prob. = 0		**10,584**	**10,211**	**373**	**317**	**10,901**
With Match Codes (DP, FP, GP, MN, NL, NN)		10,502	10,129	373	317	10,819
No Match Codes						
RSC = N or U, MOVERPER = 1		48	48	0	0	48
RSC = I, MOVERPER = 1		34	34	0	0	34
Out-movers (MOVERPER=3, RSC≠R)		**25,345**	**23,434**	**1,911**	**748**	**26,093**
Census Day Resident Prob. = 1		**19,047**	**17,578**	**1,469**	**588**	**19,635**
Matches						
M		13,939	13,128	811	0	13,939
MR		1,439	1,355	84	0	1,439
Nonmatches						
NR		2,626	2,254	372	337	2,963
NP		964	762	202	251	1,215
NC		79	79	0	0	79
Census Day Resident Prob. (0,1)		**4,408**	**4,224**	**184**	**115**	**4,523**
Matches	MU	534	488	46	0	534
Nonmatches	NU	1,410	1,272	138	115	1,525
Other						
KI		1,399	1,399	0	0	1,399
KP		1,046	1,046	0	0	1,046
P		19	19	0	0	19
Census Day Resident Prob. = 0 (DP, FP, GP, MN, NL, NN)		**1,890**	**1,632**	**258**	**45**	**1,935**
In-movers (MOVERPER=2, RSC≠R)		**36,623**	**36,623**	**0**	**0**	**36,623**
RSC = N; NN		1	1	0	0	1
RSC = I		36,622	36,622	0	0	36,622
Removed (RSC=R)		**16,203**	**16,203**	**0**	**0**	**16,203**

Table 7. E-Sample Person Counts for Missing Data & Estimation (Unweighted)

1/08/01

	Total	Non-TES People	TES People (TESWGT > 0)	TES People (TESWGT = 0)	Overall Total
Total	**704,602**	**680,566**	**24,036**	**8,298**	**712,900**
Correct Enumeration	**652,390**	**633,644**	**18,746**	**3**	**652,393**
Matches					
M	539,080	539,080	0	0	539,080
MR	9,668	9,668	0	0	9,668
Nonmatches					
CE	103,642	84,896	18,746	3	103,645
Unresolved Enumeration	**21,148**	**18,029**	**3,119**	**2**	**21,150**
Matches					
MU	897	897	0	0	897
Nonmatches					
UE	19,510	16,997	2,513	2	19,512
GU	625	19	606	0	625
Other					
P	116	116	0	0	116
Erroneous Enumeration	**31,064**	**28,893**	**2,171**	**8,293**	**39,357**
Matches					
MN	198	198	0	0	198
Nonmatches					
EE	7,428	7,081	347	0	7,428
DE	5,990	5,759	231	0	5,990
FE	2,026	1,949	77	0	2,026
GE	2,062	546	1,516	8,293	10,355
Other					
KE	13,360	13,360	0	0	13,360

ESCAP MEETING NO. 28 - 01/17/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting #28
January 17, 2001

Prepared by: Nick Birnbaum

The twenty-eighth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on January 17, 2001 at 10:30.

The agenda for the meeting was to examine the A.C.E. missing data procedures and results.

Committee Attendees:

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Other Attendees:

Kenneth Prewitt
Marvin Raines
Tommy Wright
Donna Kostanich
Raj Singh
Gregg Robinson
Donald Malec
Patrick Cantwell
Richard Griffin
Kathleen Styles
Nick Birnbaum
Sarah Brady
Carolee Bush
Annette Quinlan
Maria Urrutia

I. A.C.E. Missing Data Procedures and Results

Pat Cantwell walked the Committee through a document describing the missing data procedures and providing results of the missing data operations.

He briefly discussed the impact of missing data on the DSE results. Missing data can contribute to variance, and, if the missing data models are poorly specified, can also contribute to bias in the estimates. There are three components to missing data:

- Whole household non-interviews
- Missing post-stratification variables
- Unresolved match, residence or enumeration status.

Pat began by providing data on A.C.E. household interviews for Census Day and for A.C.E. Interview Day. It was noted the interview rates were very high. As expected, the interview rate was slightly higher for A.C.E. interview day residents than Census Day residents. The comparable interview rate for the 1990 PES fell between these levels. The Committee was pleased with the low levels of A.C.E. missing data.

Pat then discussed characteristic imputation. The five variables that are imputed when missing are those needed for post-stratification assignment. He noted the different procedures used for P and E-sample imputation. Generally, for the E-sample, the imputed value from the census edited file is used. It was noted that the characteristic imputation rates were low, as in the Dress Rehearsal. Pat also presented data showing the missing data rates for each of the five post-stratification variables and the distribution of these characteristic responses before and after imputation for the weighted P and E-samples. It was observed that the frequency of "other race only" was noticeably greater in the P-sample distribution that in the E-sample distribution. Concerns were expressed about the impact this could have on the consistency of post-stratification, potentially increasing heterogeneity or correlation bias.

Next, Pat discussed the imputation of status (resident, match, and enumeration). Even after all the follow-up activities have been completed, there remains a small fraction of the A.C.E. sample people for whom we still do not have enough information to compute the components of the dual system estimator. Their status is said to be "unresolved." The Census Bureau uses imputation cell estimation to assign probabilities for P-Sample people with unresolved match or Census-Day resident status, and for E-sample people with unresolved enumeration status. In imputation cell estimation, all P and E-sample people – both resolved and unresolved cases – are divided up into groups called imputation cells based on similar operational and demographic characteristics. Within each imputation cell, the weighted proportion of matches (or residents or correct enumerations) among the cases with resolved status is calculated, and that value is imputed for all unresolved people in the cell. This methodology was briefly contrasted with the more complicated hierarchical logistic regression model used in the 1990 PES to assign probabilities to unresolved cases.

Examining the results of imputation of status, Pat noted that the proportion of P-sample people with unresolved residence status was very low and thus, imputation procedures for these cases would appear to have a very minor effect on the estimation process. With regard to match status, the proportion of P-sample people having unresolved match status was low; again, implying only a small effect on the estimation. Finally, enumeration status was examined. The proportion of E-sample cases with unresolved enumeration status was believed to be roughly comparable to the 1990 PES rate, perhaps slightly lower. The results of the change in the imputation cells for certain categories of unresolved enumeration status cases (see Section I of the January 10, 2001 ESCAP meeting minutes) were discussed. It was noted that this change resulted in the assignment of more precise probabilities, and that the decision to make the change was supported by these results.

II. Next Meeting

The agenda for the next meeting, scheduled for January 24, 2001, is to address residual questions from today's discussion, and to examine the consistency of post-stratification variables.

ESCAP MEETING NO. 29 - 01/24/01

AGENDA

Kathleen P Porter
01/23/2001 11:31 AM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc:

Subject: Agenda for 1/24 ESCAP meeting

The agenda for the January 24 ESCAP Meeting scheduled from 10:30-12:00 in Rm. 2412/3 is as follows:

1. Missing Data Questions, if any - Pat Cantwell

2. Census Race Classifications - Greg Robinson

3. Consistency of Poststratification Variances - Jim Farber

4. Synthetic Error Methodology - Rick Griffin
(please remember to bring information handed out at last meeting)

ESCAP MEETING NO. 29 - 01/24/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Consistency of A.C.E. Post-stratification Variables

- Consistency is the agreement of housing and person post-stratification variables for matched people in the P sample and E sample. For example, a person who reports her age as 28 in the P sample and 29 in the E sample is consistent because that person is in the same group of the Age/Sex post-stratification variable.

- We assess consistency for three post-stratification variables: Tenure, Age/Sex, and Race/Hispanic Origin domain. Tables A-1, A-2, and A-3 show the consistency results.

- We can not quantify at this point how the level of inconsistency affects the dual system estimates (DSEs). However, inconsistency generally does not affect the DSEs as much as other types of estimates, such as those based on demographic analysis (DA), because the DSEs are based on ratios. Misclassified people are taken out of the numerator and denominator of the DSE, and thus the coverage correction factor does not change significantly. In additive estimates, such as DA, misclassification can have a larger effect.

- Inconsistency is not a concern if people are misclassified into two post-strata with similar coverage properties. If the rates of inconsistency are large and people are misclassified into post-strata with different coverage properties, then heterogeneity bias can be introduced into the DSEs. We can not yet measure the heterogeneity created by inconsistent post-stratification.

- Overall, the rates of inconsistency for matched people are about 5 percent or lower for the post-stratification variables.

- Imputation significantly increases the level of inconsistency. For Age/Sex, almost 50 percent of the inconsistency in the A.C.E. is caused by the characteristic imputation procedure.

Table A-1: Consistency of A.C.E. Post-Stratification Variables: Tenure

Total Matched Cases		E Sample Owner	E Sample Non-Owner	Total	% Inconsistent
P Sample	Owner	369,965	11,632	381,597	3.05%
	Non-Owner	14,129	153,022	167,151	8.45%
	Total	384,094	164,654	548,748	
	% Inconsistent	3.68%	7.06%		4.69%
Non-Imputed Cases					
P Sample	Owner	354,026	8,670	362,696	2.39%
	Non-Owner	10,826	146,425	157,251	6.88%
	Total	364,852	155,095	519,947	
	% Inconsistent	2.97%	5.59%		3.75%

Table A-2: Consistency of A.C.E. Post-Stratification Variables: Age/Sex

Total Matched Cases		E Sample: 0 - 17	18 - 29 M	18 - 29 F	30 - 49 M	30 - 49 F	50+ M	50 + F	Total	% Inconsistent
P Sample	0 - 17	143,847	738	569	359	326	190	245	146,274	1.66%
	18 - 29 M	687	36,184	991	1,038	43	398	23	39,364	8.08%
	18 - 29 F	573	1,327	38,227	43	1,052	11	458	41,691	8.31%
	30 - 49 M	225	1,108	51	77,227	1,870	1,635	57	82,173	6.02%
	30 - 49 F	196	57	974	2,828	83,636	75	1,889	89,655	6.71%
	50 + M	77	229	15	1,431	62	63,719	1,676	67,209	5.19%
	50 + F	107	13	224	62	1,479	2,529	77,968	82,382	5.36%
	Total	145,712	39,656	41,051	82,988	88,468	68,557	82,316	548,748	
	% Inconsistent	1.28%	8.76%	6.88%	6.94%	5.46%	7.06%	5.28%		5.09%
Non-Imputed Matched Cases										
P Sample	0 - 17	135,581	268	202	101	123	76	104	136,455	0.64%
	18 - 29 M	331	34,708	694	447	17	31	3	36,231	4.20%
	18 - 29 F	278	1,012	36,906	17	444	3	23	38,683	4.59%
	30 - 49 M	50	608	26	74,049	1,250	583	14	76,580	3.31%
	30 - 49 F	65	21	557	2,163	80,105	20	691	83,622	4.21%
	50 + M	11	35	3	579	20	61,025	1,117	62,790	2.81%
	50 + F	23	4	25	13	554	1,891	74,221	76,731	3.27%
	Total	136,339	36,656	38,413	77,369	82,513	63,629	76,173	511,092	
	% Inconsistent	0.56%	5.31%	3.92%	4.29%	2.92%	4.09%	2.56%		2.84%

Table A-3: Consistency of A.C.E. Post-Stratification Variables: Race/Hispanic Origin Domains

Total Matched Cases		E Sample: Amer Ind on Res	Amer Ind off Res	Hispanic	Black	Native Hawaiian	Asian	Non-Hisp White	Total	% Incon.
P Sample	Amer Ind on Res	11,007	0	34	12	0	0	118	11,171	1.47%
	Amer Ind off Res	0	2,211	59	104	0	30	785	3,189	30.67%
	Hispanic	44	136	67,888	608	42	266	3,983	72,967	6.96%
	Black	10	119	494	65,566	6	117	1,409	67,721	3.18%
	Native Hawaiian	0	3	31	19	1,669	203	177	2,102	20.60%
	Asian	1	31	107	99	143	19,648	1,051	21,080	6.79%
	Non-Hisp White	107	939	5,019	2,568	183	2,093	359,609	370,518	2.94%
	Total	11,169	3,439	73,632	68,976	2,043	22,357	367,132	548,748	
	% Inconsistent	1.45%	35.71%	7.80%	4.94%	18.31%	12.12%	2.05%		3.85%
Non-Imputed Matched Cases										
P Sample	Amer Ind on Res	10,484	0	24	10	0	0	103	10,621	1.29%
	Amer Ind off Res	0	2,027	48	84	0	25	703	2,887	29.79%
	Hispanic	28	84	54,053	401	34	177	2,991	57,768	6.43%
	Black	10	94	349	59,341	5	80	1,068	60,947	2.64%
	Native Hawaiian	0	3	15	16	1,550	178	147	1,909	18.81%
	Asian	1	15	72	68	110	18,015	737	19,018	5.27%
	Non-Hisp White	93	844	3,514	2,063	141	1,718	343,314	351,687	2.38%
	Total	10,616	3,067	58,075	61,983	1,840	20,193	349,063	504,837	
	% Inconsistent	1.24%	33.91%	6.93%	4.26%	15.76%	10.79%	1.65%		3.18%

Note: The race/Hispanic origin groups in the table correspond to the race/origin domains assigned in A.C.E. post-stratification. See Appendix 2 of the attached memorandum for more information on the definition and assignment of these domains.

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

DRAFT - January 24, 2001

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-???

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Donna Kostanich
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: James Farber
Sample Design Team

Subject: Accuracy and Coverage Evaluation Survey: Consistency of Post-Stratification Variables

The attached document was prepared at your request to assist the Executive Steering Committee on A.C.E. Policy in its recommendation regarding the release of the statistically corrected data or data without statistical correction.

This report focuses on the consistency of post-stratification variables between the P sample and E sample. The analysis is limited to P-sample and E-sample cases that matched following person matching.

Accuracy and Coverage Evaluation: Consistency of Post-Stratification Variables

James Farber
U.S. Census Bureau

Introduction

Executive Summary

Results

References

Appendices

Tables

Accuracy and Coverage Evaluation 2000: Consistency of Post-Stratification Variables

prepared by James Farber

Introduction

The Accuracy and Coverage Evaluation (A.C.E.) consists of two independent samples. The first is a sample of the population in selected A.C.E. sample areas, known as the Population or P sample. Matching these people to census records provides an estimate of the proportion of the population missed in the census. The second is a sample of the census enumerations in the same A.C.E. sample areas, known as the Enumeration or E sample. An estimate of the proportion of correctly enumerated census people can be determined using the results of matching the P sample to the census, checking for duplication among the census records, and re-interviewing when needed to determine the correct inclusion of each E-sample person. Together, these estimates measure the net omissions or erroneous enumerations in the census.

The A.C.E. includes dual system estimates for up to 448 post-strata for the 50 states and the District of Columbia (Haines, 2000). Each P-sample person and E-sample person is assigned to a post-stratum based on the values of race, Hispanic origin, age, sex, and tenure. Ideally, a P-sample person who matches to an E-sample person will have consistent values for those post-stratification variables, but in reality this may not occur. If a matched person does not have consistent characteristics in his P-sample and E-sample records, then that single person could be misclassified into two different post-strata when estimating the net proportions of people missed or correctly enumerated in the census.

The purpose of this report is to get an indication of the consistency of the post-stratification variables between P sample and E sample. Persistent differences in the classification of people in the census and the A.C.E. may introduce a potential bias into the coverage estimates. This bias is sometimes referred to as classification error.

One reason the two systems may differ is when a person has an unknown post-stratification variable that is filled in through characteristic imputation. Both the P sample and E sample are subject to characteristic imputation. Appendix 1 gives the criteria for determining if a characteristic has undergone imputation. There are separate criteria for the P sample and the E sample. For this analysis, if a characteristic is imputed for either the P sample or the E sample, then the entire case is considered imputed for that characteristic.

This report distinguishes between imputed and non-imputed characteristics. Some tables display results for total cases and non-imputed cases. The corresponding results for imputed cases are the difference between the total and non-imputed. This decomposition clarifies the source of inconsistency. For imputed cases, inconsistency is often attributable to the characteristic imputation procedure. For non-imputed cases, inconsistencies arise due to inconsistent reporting, which has many possible causes including the data collection mode, time lag from reference day, proxy responses, or data capture difficulties. This report does not explore these reasons for misclassification.

This report also excludes adjustments for P-sample matches to E-sample duplicates. In the person matching process, a P-sample case may possibly match to an E-sample case later found to be a duplicate of another E-sample case or a non-E-sample census record. In this situation, we use only the matched P-sample and E-sample records to assess consistency.

A similar analysis on consistency was conducted for the 1998 Census 2000 Dress Rehearsal and the 1995 Census Test (see Salganik, 1999 and Petroni, 1996A and 1996B). An analysis on the 1990 Post-Enumeration Survey (PES) data is desirable but information linking PES P-sample and E-sample records is not readily available, making such an analysis impossible at this time.

Executive Summary

What are the overall rates of inconsistency?

Overall, the rates of inconsistency for the matched cases are about 5 percent or lower for all of the post-stratification variables. We do not know at this point how this level of inconsistency affects the A.C.E. estimates.

What are the overall rates of non-balanced cases?

The overall rates of non-balanced cases are about 0.6% or lower for all of the post-stratification variables.

How does imputation affect inconsistency and non-balance?

- Imputed cases have significantly greater rates of inconsistency than non-imputed cases for all of the post-stratification variables.

- Imputed cases have significantly greater rates of non-balanced cases than non-imputed cases for all post-stratification variables.

Results

Measuring consistency

A variable is defined as consistent when the information collected for a person is the same in the P sample and E sample or results in the classification of the person to the same level of the post-stratification variable. For example, a person who reports her age as 28 in the P sample and 27 in the E sample would be classified in the 18 - 29 group of the age post-stratification variable, and thus that person's age is consistent even though it does not match exactly.

To measure the consistency of post-stratification variables, we are limited to looking at P-sample cases that matched to an E-sample case following the A.C.E. person matching operation. This means that P-sample cases that matched to a census enumeration not in the E sample are excluded from this analysis. Including such cases is technically feasible but difficult in practice, and the gain in assessing consistency would likely be minimal since there are relatively few of these cases compared to the number of matches. We will be able to detect any classification error problem using the large amount of readily available data from matched cases.

This report also looks at whether misclassifications are balanced. Inconsistencies that occur randomly and are balanced are of less concern than systematic switching from one group to another, an imbalanced scenario. Note, though, that even imbalanced inconsistency is a concern only when the matched person's two different post-strata have significantly different coverage properties. If the coverage correction factors for the two post-strata are very similar, then the misclassification has no practical effect on the A.C.E. population estimates. Classification error is a function of not only the amount of inconsistency but also the differences in coverage rates among the post-strata. This report does not include an analysis of the coverage rates of the post-strata in the various combinations of inconsistent cases.

This study may under-report the amount of inconsistency because the data include only matched cases. The non-matched people may be more inconsistent simply because they cannot be matched. Use caution when drawing conclusions about the entire population based on the consistency of only the matched people. Ideally, this report would include the non-matched cases to obtain an overall measure of inconsistency, but this is not possible. Assessing consistency requires that the P-sample and E-sample information be linked. Non-matched people do not have that link, and thus can not be studied.

Post-stratification variables

The post-stratification variables considered in this analysis are tenure, age/sex, and race/Hispanic origin domain. All other post-stratification variables are geographically assigned variables that by definition are consistent between the P and E samples. A person is consistent if their P-sample and E-sample responses are in the same group of each post-stratification variable, as listed below:

Tenure

- Owner
- Non-owner

Age/Sex

- Under 18
- 18 - 29 Male
- 18 - 29 Female
- 30 - 49 Male
- 30 - 49 Female
- 50 + Male
- 50 + Female

Race/Hispanic Origin Domain (see Appendix 2 for more detail on these seven domains)

- Domain 1 American Indian or Alaska Native on reservations
- Domain 2 American Indian or Alaska Native off reservations
- Domain 3 Hispanic
- Domain 4 Non-Hispanic Black
- Domain 5 Native Hawaiian or Pacific Islander
- Domain 6 Non-Hispanic Asian
- Domain 7 Non-Hispanic White or "Some Other Race"

Table 1 below summarizes the consistency for each of these three post-stratification variables by imputation status. See Tables A-1, A-2, and A-3 in Attachment A for more detailed results. The total number of matched cases is the same for all three variables, but their distributions differ by imputation status.

These tables also show non-balanced inconsistent cases, which are the absolute difference of the inconsistent cases. For example, Table A-1 shows of the total matched cases there are

- 11,632 people who are owners in the P sample but non-owners in the E sample
- 14,129 people who are non-owners in the P sample but owners in the E sample

Thus for tenure there are

- 11,632 + 14,129 = 25,761 inconsistent cases (about 4.7% of the total matches)
- 14,129 - 11,632 = 2,497 non-balanced cases (about 0.5% of the total matches)

Table 1: Consistency of Matching P-Sample and E-Sample Post-Stratification Variables
(source: A.C.E. missing data P-sample and E-sample person output files)

Variable	Total Cases	Consistent Cases	Inconsistent		Non-Balanced	
			Cases	Percent	Cases	Percent
Tenure	548,748	522,987	25,761	4.69%	2,497	0.46%
Non-Imputed	519,947	500,451	19,496	3.75%	2,156	0.41%
Imputed	28,801	22,536	6,265	21.75%	341	1.18%
Age/Sex	548,748	520,808	27,940	5.09%	568	0.10%
Non-Imputed	511,092	496,595	14,497	2.84%	2,041	0.40%
Imputed	37,656	24,213	13,443	35.70%	1,473	3.91%
Race/Origin Domain	548,748	527,598	21,150	3.85%	3,142	0.57%
Non-Imputed	504,837	488,784	16,053	3.18%	2,423	0.48%
Imputed	43,911	38,814	5,097	11.61%	719	1.64%

448 Post-strata

Tables B-1 through B-64 in Attachment B show consistency results for each of the 64 major post-stratum groups by the 7 age/sex groups.

References

Haines, Dawn, "Accuracy and Coverage Evaluation Survey: Final Post-stratification Plan for Dual System Estimation," DSSD Census 2000 Procedures and Operations Memorandum Series Q-24, April 19, 2000.

Petroni, Rita, "Disagreement of Characteristics Between R-Sample and Census Linked Cases for Oakland - Preliminary Findings," Internal Census Bureau Memorandum, March 11, 1996.

Petroni, Rita, "Disagreement of Characteristics Between R-Sample and Census Linked Cases for Oakland - More Findings," Internal Census Bureau Memorandum, May 7, 1996.

Salganik, Matt, "Accuracy and Coverage Evaluation Survey: Consistency of Potential Poststratification Variables," DSSD Census 2000 Procedures and Operations Memorandum Series Q-10, September 23, 1999.

Table A-1: Consistency of Post-Stratification Variables: Tenure

Total Matched Cases		E Sample Owner	E Sample Non-Owner	Total	% Inconsistent
P Sample	Owner	369,965	11,632	381,597	3.05%
	Non-Owner	14,129	153,022	167,151	8.45%
	Total	384,094	164,654	548,748	
	% Inconsistent	3.68%	7.06%		4.69%
Non-Imputed Cases					
P Sample	Owner	354,026	8,670	362,696	2.39%
	Non-Owner	10,826	146,425	157,251	6.88%
	Total	364,852	155,095	519,947	
	% Inconsistent	2.97%	5.59%		3.75%

Table A-2: Consistency of Post-Stratification Variables: Age/Sex

Total Matched Cases		E Sample 0 - 17	18 - 29 M	18 - 29 F	30 - 49 M	30 - 49 F	50+ M	50 + F	Total	% Inconsistent
P Sample	0 - 17	143,847	738	569	359	326	190	245	146,274	1.66%
	18 - 29 M	687	36,184	991	1,038	43	398	23	39,364	8.08%
	18 - 29 F	573	1,327	38,227	43	1,052	11	458	41,691	8.31%
	30 - 49 M	225	1,108	51	77,227	1,870	1,635	57	82,173	6.02%
	30 - 49 F	196	57	974	2,828	83,636	75	1,889	89,655	6.71%
	50 + M	77	229	15	1,431	62	63,719	1,676	67,209	5.19%
	50 + F	107	13	224	62	1,479	2,529	77,968	82,382	5.36%
	Total	145,712	39,656	41,051	82,988	88,468	68,557	82,316	548,748	
	% Inconsistent	1.28%	8.76%	6.88%	6.94%	5.46%	7.06%	5.28%		5.09%

Non-Imputed Matched Cases		0 - 17	18 - 29 M	18 - 29 F	30 - 49 M	30 - 49 F	50+ M	50 + F	Total	% Inconsistent
P Sample	0 - 17	135,581	268	202	101	123	76	104	136,455	0.64%
	18 - 29 M	331	34,708	694	447	17	31	3	36,231	4.20%
	18 - 29 F	278	1,012	36,906	17	444	3	23	38,683	4.59%
	30 - 49 M	50	608	26	74,049	1,250	583	14	76,580	3.31%
	30 - 49 F	65	21	557	2,163	80,105	20	691	83,622	4.21%
	50 + M	11	35	3	579	20	61,025	1,117	62,790	2.81%
	50 + F	23	4	25	13	554	1,891	74,221	76,731	3.27%
	Total	136,339	36,656	38,413	77,369	82,513	63,629	76,173	511,092	
	% Inconsistent	0.56%	5.31%	3.92%	4.29%	2.92%	4.09%	2.56%		2.84%

Table A-3: Consistency of Post-Stratification Variables: Race/Hispanic Origin Domains

Total Matched Cases		E Sample								
		Domain 1	Domain 2	Domain 3	Domain 4	Domain 5	Domain 6	Domain 7	Total	% Incon.
P Sample	Domain 1	11,007	0	34	12	0	0	118	11,171	1.47%
	Domain 2	0	2,211	59	104	0	30	785	3,189	30.67%
	Domain 3	44	136	67,888	608	42	266	3,983	72,967	6.96%
	Domain 4	10	119	494	65,566	6	117	1,409	67,721	3.18%
	Domain 5	0	3	31	19	1,669	203	177	2,102	20.60%
	Domain 6	1	31	107	99	143	19,648	1,051	21,080	6.79%
	Domain 7	107	939	5,019	2,568	183	2,093	359,609	370,518	2.94%
	Total	11,169	3,439	73,632	68,976	2,043	22,357	367,132	548,748	
	% Inconsistent	1.45%	35.71%	7.80%	4.94%	18.31%	12.12%	2.05%		3.85%
Non-Imputed Matched Cases										
P Sample	Domain 1	10,484	0	24	10	0	0	103	10,621	1.29%
	Domain 2	0	2,027	48	84	0	25	703	2,887	29.79%
	Domain 3	28	84	54,053	401	34	177	2,991	57,768	6.43%
	Domain 4	10	94	349	59,341	5	80	1,068	60,947	2.64%
	Domain 5	0	3	15	16	1,550	178	147	1,909	18.81%
	Domain 6	1	15	72	68	110	18,015	737	19,018	5.27%
	Domain 7	93	844	3,514	2,063	141	1,718	343,314	351,687	2.38%
	Total	10,616	3,067	58,075	61,983	1,840	20,193	349,063	504,837	
	% Inconsistent	1.24%	33.91%	6.93%	4.26%	15.76%	10.79%	1.65%		3.18%

Tables B-1 through B-64: Consistency of P and E Samples by Post-Stratum Group

Post-Stratum Definition	M & F 0 - 17	Males 18 - 29	Females 18 - 29	Males 30 - 49	Females 30 - 49	Males 50+	Females 50+	Total
Total P & E Sample Matches								
Total Cases								
Total Consistent Persons								
Proportion Consistent								
Total P & E Sample Matches - Non-Imputed								
Total Cases								
Total Consistent Persons								
Proportion Consistent								
Total P & E Sample Matches - Imputed								
Total cases								
Total Consistent Persons								
Proportion Consistent								

Appendices

Appendix 1. Characteristic Imputation

Appendix 2. Race/Hispanic Origin Domains

Characteristic Imputation

P Sample

For dual system estimation, P-sample records are imputed because of either an edit failure or a missing value Table 1.1 identifies what is considered an imputed value for the variables needed to assign P-sample records to the post-strata. The values in the table are the imputation flag values for these variables. The flags can be found on the Person Dual System Estimation P-sample Output Person File.

Table 1.1: Identifying Imputed Values for the P Sample

P-sample Characteristic	Reported Values	Imputed Values
Age	1 = No imputation	2 = Imputation because of edit failure 3 = Imputation because of missing value
Race	1 = No imputation	2 = Imputation because of edit failure 3 = Imputation because of missing value
Hispanic Origin	1 = No imputation	2 = Imputation because of edit failure 3 = Imputation because of missing value
Sex	1 = No imputation	2 = Imputation because of edit failure 3 = Imputation because of missing value
Tenure	1 = No imputation	2 = Imputation because of edit failure 3 = Imputation because of missing value

E Sample

For dual system estimation, E-sample records are imputed using the results of the census imputation procedure. We match each E-sample housing unit or person to the Hundred Percent Census Edited File (HCEF), and fill in missing characteristics in the E-sample records with the corresponding imputed values in the HCEF records. Table 1.2 identifies what is considered an imputed value for the variables needed to assign E-sample records to a post-stratum. The values in the table are the HCEF allocation flag values for these variables. The flags can be found on the Person Dual System Estimation E-sample Output Person File.

Table 1.2: Identifying Imputed Values for the E Sample

E-sample Characteristic	Reported Values	Imputed Values
Age	0 = Both Consistent 1 = Age Only 2 = Date of birth only	3 = Inconsistent age and date of birth 4 = Allocated from hot deck 9 = E-sample person did not match to the HCEF
Race	0 = As reported	1 = Code changed through consistency edit 3 = Assigned from race response to Hispanic origin question 4 = Allocated from within household 5 = Allocated from hot deck 9 = E-sample person did not match to the HCEF
Hispanic Origin	0 = 1 reported origin 2 = Multiple response given a unique Hispanic or Non-Hispanic code	3 = Assigned Hispanic Origin from race code 4 = Allocated from within household 5 = Allocated from hot deck (surname used) 6 = Allocated from hot deck (surname not used) 9 = E-sample person did not match to the HCEF
Sex	0 = As reported	1 = From first name 2 = Edited 4 = Allocated from hot deck 5 = Allocated from consistency check 9 = E-sample person did not match to the HCEF
Tenure	0 = As reported	1 = Assigned by consistency check 4 = Allocated from hot deck 9 = E-sample person did not match to the HCEF

Race/Hispanic Origin Domains

The Race/Hispanic origin domain assignment is hierarchical. See Haines (2000) for more detail.

Domain 1 (American Indian or Alaska Native on reservations) includes:

- Any person living on a reservation indicating American Indian or Alaska Native either as their single race or as one of many races, regardless of their Hispanic origin.

Domain 2 (American Indian or Alaska Native off reservations) includes:

- Any person living in Indian Country[1] but not on a reservation who indicates American Indian or Alaska Native either as their single race or as one of many races, regardless of their Hispanic origin.

- Any non-Hispanic person not living in Indian Country who indicates American Indian or Alaska Native as their single race.

Domain 3 (Hispanic) includes:

- All Hispanic persons who are not included in Domains 1 or 2.

- All Hispanic persons who self-identify with three or more races (excluding American Indian or Alaska Native in Indian Country).

- All Hispanic persons who do not live in the state of Hawaii who classify themselves as Native Hawaiian or Pacific Islander, regardless of whether they identify with a single or multiple race.

[1] Indian Country is land considered (either wholly or partially) on an American Indian reservation/trust land, Tribal Jurisdiction Statistical Area, Tribal Designated Statistical Area, or Alaska Native Village Statistical Area. For Census 2000, Tribal Jurisdiction Statistical Area has been formally renamed as Oklahoma Tribal Statistical Area.

Domain 4 (Non-Hispanic Black) includes:

- Any non-Hispanic person who indicates Black as their only race.
- Any person identifying with a combination of Black and American Indian or Alaska Native not in Indian Country.
- Any person who indicates Black and another single race group (Native Hawaiian or Pacific Islander, Asian, White, or "Some other race").
- All Non-Hispanic Black persons who do not live in the state of Hawaii who classify themselves as Native Hawaiian or Pacific Islander.

Domain 5 (Native Hawaiian or Pacific Islander) includes:

- Any non-Hispanic person indicating the single race Native Hawaiian or Pacific Islander.
- Any non-Hispanic person who identifies with the race combination of Native Hawaiian or Pacific Islander and American Indian or Alaska Native not in Indian Country.
- Any non-Hispanic person who identifies with the race combination of Native Hawaiian or Pacific Islander and Asian.
- All persons living in the state of Hawaii who classify themselves as Native Hawaiian or Pacific Islander, regardless of their Hispanic origin and whether they identify with a single or multiple race.

Domain 6 (Non-Hispanic Asian) includes:

- Any non-Hispanic person indicating Asian as their single race.
- Any person who self-identifies with Asian and American Indian or Alaska Native not in Indian Country.

Domain 7 (Non-Hispanic White or "Some other race") includes:

- Any Non-Hispanic person indicating White or "Some other race" as their single race.

- Any Non-Hispanic person who self-identifies with both American Indian or Alaska Native not in Indian Country and White or "Some other race."

- Any person who self-identifies with Asian and White or Asian and "Some other race."

- Any non-Hispanic person who self-identifies with three or more races (excluding American Indian or Alaska Native in Indian Country).

- Any Non-Hispanic White or Non-Hispanic "Some other race" person who classifies themselves as Native Hawaiian or Pacific Islander but does not live in Hawaii, regardless of whether they identify with other races.

ESCAP MEETING NO. 29 - 01/24/01

MINUTES

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 29

January 24, 2001

Prepared by: Maria Urrutia and Sarah Brady

The twenty-ninth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on January 24, 2001 at 10:30. The agenda for the meeting was to answer any questions about the missing data presentation from the previous ESCAP meeting and to present data illustrating the consistency of the poststratification variables.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
Ruth Ann Killion
Howard Hogan
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines	William Bell
Tommy Wright	Donald Malec
Gregg Robinson	Louisa Miller
Raj Singh	Nick Birnbaum
Donna Kostanich	Kathleen Styles
Patrick Cantwell	Maria Urrutia
Richard Griffin	Annette Quinlan
Deborah Fenstermaker	Sarah Brady
James Farber	Carolee Bush

I. Questions on Results from Missing Data

John Thompson began the meeting by asking the Committee members if there were any questions about the missing data results that were presented at the last ESCAP meeting. The Committee did not have any questions.

II. Consistency of Post-stratification Variables

Howard Hogan introduced the topic of consistency of the A.C.E. post-stratification variables. Consistency is the agreement of the responses for housing and person post-stratification variables for people in the P-sample and E-sample. For instance, a person who reports his age as 28 in the P-sample and 29 in the E-sample is consistent for the age/sex post-stratification variable, since both of those ages are in the same group of the age/sex post-stratum. We can study this response consistency for the matched persons between the P and E-samples.

Jim Farber presented an analysis of the consistency of matched cases for three of the post-stratification variables: Tenure, Age/Sex, and Race/Hispanic Origin domain. Overall, the rates of consistency for matched people look good for these post-stratification variables; the rates of inconsistency are about 5 percent or lower. However, there are two notable exceptions in the Race/Hispanic Origin post-stratification variables: the American Indian off-Reservations domain and the Native Hawaiian domain. These distinct domains were developed under the advisement of the Census Advisory Committee on these populations because creating the Native Hawaiian and American Indian off-Reservations domains recognizes that these groups have unique coverage properties. Without these distinct post-stratification domains, the alternative would be to define a post-stratification that would include these groups with the Non-Hispanic Whites and “Some other race” domain. The coverage correction factors for these groups would be averaged with a larger group – Non-Hispanic Whites and “Some other race.” Therefore, the measure of their undercoverage would likely be understated in the estimates. The consequence of a higher rate of inconsistency for American Indians off-Reservations and Native Hawaiians is that the measurement of undercoverage for these groups will be somewhat reduced, but still in the right direction.

The data also illustrated that imputation creates additional inconsistency. By necessity (to maintain independence) imputation is performed independently for the E and P-samples. Imputation for the E-sample can draw upon data from the entire census, whereas, the only source of data for P-sample imputation is the P-sample. Consequently, the same degree of geographic proximity is not attained for the E and P-samples. For the nearest neighbor imputation method, the nearest neighbor used as a donor in the P-sample may be in another block cluster. In contrast, the nearest neighbor in the E-sample may be in a neighboring block. The distance between donors in the P-sample is usually greater than the distance between donors in the E-sample. Due to these inherent limitations to P-sample imputation, inconsistency between the E and P-samples increases for imputed cases.

Finally, it was noted that the response inconsistency does not affect the DSE as much as it affects demographic analysis. The primary reason is that we are able to form post-strata that are reasonably consistent, due to the collection methodologies of the A.C.E. and the census. In contrast, demographic analysis is based on a single race reporting system, resulting in inconsistency with the census multiple race reporting system.

In conclusion, inconsistency will introduce additional heterogeneity within post-strata which will be included in the analysis of the synthetic assumption. The inconsistency within the two post-stratum groups described above will result in estimates of undercount which are in the right direction, but do not fully measure the undercount. However, the estimates are most likely to be closer to the true undercount than alternatives that could be obtained by alternative post-stratification designs.

II. Next Meeting

The agenda for the next meeting, scheduled for January 26, 2001, is to examine results from the A.C.E. Dual Systems Estimation and their standard errors.

ESCAP MEETING NO. 30 - 01/26/01

AGENDA

There was no agenda developed or used for the January 26, 2001 meeting.

ESCAP MEETING NO. 30 - 01/26/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Preliminary Dual System Estimation Results and Their Reliability

- Dual system estimates are based on Census 2000 and the A.C.E.

$$DSE = DD \times \frac{CE}{N_e} \times \frac{N_p}{M}$$

$$CCF = \frac{DSE}{C}$$

- All 2000 net undercount estimates are for the household population. The 1990 net undercount estimates included some Group Quarters population.

$$UC = \frac{DSE - C}{DSE} \times 100$$

- Race/origin estimation domains used in 1990 and 2000 are different. For large groups, they are reasonably comparable.

Table A-1: 2000 A.C.E. Coverage Estimates

Characteristic	Net Undercount (%)	Coverage Correction Factor	Data Defined (%)	Correct Enumeration Rate (%)	Match Rate (%)	CV(CCF) (%)
Total	1 18	1.0119	97.07	95 28	91.59	0.14
Race/Origin						
Non-Hispanic White	0.67	1.0068	97.70	95.90	93.15	0.14
Non-Hispanic Black	2.17	1.0221	95.65	92.73	86.93	0.35
Hispanic	2.85	1.0294	95.21	94.46	87.48	0.40
Hawaiian or Pacific Isl	4.60	1.0483	95.41	93.05	84.91	2.90
Non-Hispanic Asian	0.96	1.0097	96.49	94.57	90.43	0.65
AI On Reservation	4.74	1.0498	94.13	95.81	85.99	1.25
AI Off Reservation	3.28	1.0339	96.24	93.97	87.86	1.38
Tenure						
Owner	0.44	1.0045	97.61	96.41	93.80	0.14
Non-Owner	2.75	1.0283	95.90	92.69	86.57	0.27
Age/Sex						
0-17	1.54	1.0157	96.00	95.94	90.84	0.19
18-29 Male	3.77	1.0391	96.35	92.90	86.49	0.34
18-29 Female	2.23	1.0228	96.54	93.62	88.55	0.30
30-49 Male	1.86	1 0190	97.47	95.22	91.24	0.20
30-49 Female	0.96	1.0097	97.63	96.00	92.92	0.17
50+ Male	-0.25	0.9975	97.96	95.35	93.69	0.18
50+ Female	-0.79	0.9922	97.92	95.52	94.30	0.17

Net Undercount is for household population.

- Overall net undercount is 1.18 percent, or about 3.3 million people. In 1990, the overall net undercount was 1.61 percent, or approximately 4 million people.
- A differential net undercount exists for the 7 race/origin estimation domains. The net undercount ranges from 0.67 percent for the Non-Hispanic White or Other domain to 4.74 percent for the Reservation Indians.
- For the 1990 census, the net undercount ranged from 0.68 percent for Non-Hispanic White or Others to 12.22 percent for Reservation Indians.
- The net undercount rates for the Non-Hispanic Black and Hispanic domains is 2.17 and 2.85 percent, respectively. In 1990, the corresponding rates were 4.57 and 4.99 percent. There is roughly a 50 percent reduction in the undercount rate for these two domains.
- The coverage of children improved. In 1990, their net undercount percent was 3.18 percent. This figure reduced to 1.54 percent for 2000.

Table A-2: 2000 A.C.E. Net Undercount Estimate Distribution

Characteristic	Net Undercount (× 1000)	Distribution (%)
Total	3262	100
Race/Origin		
Non-Hispanic White	1302	40
AI Off Reservation	53	2
Non-Hispanic Black	741	23
Hispanic	1014	31
Hawaiian or Pac. Islander	28	1
Non-Hispanic Asian	96	3
AI On Reservation	27	1
Tenure		
Owner	840	26
Non-Owner	2422	74
Age/Sex		
0 - 17	1127	35
18 - 29 Male	845	26
18 - 29 Female	492	15
30 - 49 Male	784	24
30 - 49 Female	414	13
50+ Male	-83	-3
50+ Female	-318	-10

- Non-Hispanic Whites account for 40 percent of the net undercount in 2000 as compared to 32 percent in 1990. American Indians Off Reservations were included in the Non-Hispanic White/Other group in 1990.

- Blacks account for less of the net undercount in 2000 (35 percent in 1990; 23 percent in 2000). The Hispanic domain remained at similar levels (28 percent in 1990; 31 percent in 2000).

- Owners account for 26 percent of the net undercount in 2000. In comparison, Owners only accounted for 2 percent of the net undercount in 1990.

- Children account for approximately 35 percent of the net undercount in 2000, compared to 52 percent in 1990.

Table A-3: 2000 A.C.E. Race/Origin by Tenure Coverage Estimates

Race/Origin by Tenure	Net Undercount (%)	Coverage Correction Factor	Data Defined (%)	Correct Enumeration Rate (%)	Match Rate (%)	CV(CCF) (%)
Non-Hispanic White	0.67	1.0068	97.70	95.90	93.15	0.14
Owner	0.30	1.0030	98.07	96.70	94.60	0.14
Non-Owner	1.85	1.0188	96.53	93.20	88.37	0.32
Non-Hispanic Black	2.17	1.0221	95.65	92.73	86.93	0.35
Owner	0.68	1.0068	96 19	94.25	90.14	0.46
Non-Owner	3.58	1.0371	95.12	91.16	83.67	0.50
Hispanic	2.85	1.0294	95.21	94.46	87.48	0.40
Owner	1.25	1.0126	95.39	96.25	90.79	0.45
Non-Owner	4.32	1.0452	95.04	92.79	84.48	0.58
Native Hawaiian or PI	4.60	1.0483	95.41	93.05	84.91	2.90
Owner	2.71	1.0278	95.51	93.79	87.36	3.94
Non-Owner	6.58	1.0704	95.30	92.33	82.39	4.36
Non-Hispanic Asian	0 96	1.0097	96.49	94.57	90.43	0.65
Owner	0.55	1.0055	96.87	95.84	92.34	0.87
Non-Owner	1.58	1.0161	95.90	92.45	87.33	1.00
AI On Reservation	4.74	1.0498	94.13	95.81	85.99	1.25
Owner	5.04	1.0531	94 00	95.65	85.43	1.52
Non-Owner	4.10	1.0428	94.41	96.15	87.08	1.48
AI Off Reservation	3.28	1.0339	96.24	93.97	87.86	1.38
Owner	1.60	1.0163	96.49	94.55	90.19	1.98
Non-Owner	5.57	1.0590	95.88	93.16	84.65	2.14
Total	1.18	1.0119	97.07	95.28	91.59	0.14
Owner	0.44	1.0045	97.61	96.41	93.80	0.14
Non-Owner	2.75	1.0283	95.90	92.69	86.57	0.27

Net Undercount is for household population.

- Owners typically have a smaller net undercount rate than Non-Owners. One exception occurs for Reservation Indians where Owners have an undercount rate of 5.04 percent while Non-Owners have a 4 10 percent undercount rate.

Table A-4: 2000 A.C.E. Tenure by Age/Sex Coverage Estimates

Tenure by Age/Sex	Net Undercount (%)	Coverage Correction Factor	Data Defined (%)	Correct Enumeration Rate (%)	Match Rate (%)	CV(CCF) (%)
Owner	0.44	1.0045	97.61	96.41	93.80	0.14
0-17	0.93 vs 2.73	1.0094	96.63	97.03	93.03	0.20
18-29 Male	0.79	1.0080	96.78	94.11	90.64	0.36
18-29 Female	0.74	1.0074	96.90	94.69	91.31	0.35
30-49 Male	1.10	1.0112	97.97	96.65	93.74	0.21
30-49 Female	0.58	1.0058	98.08	97.02	94.68	0.18
50+ Male	-0.33	0.9967	98.28	96.16	94.86	0.18
50+ Female	-0.53	0.9947	98.21	96.23	95.03	0.16
Non-Owner	2.75	1.0283	95.90	92.69	86.57	0.27
0-17	2.73	1.0280	94.76	93.64	86.39	0.39
18-29 Male	6.50	1.0695	95.93	91.65	82.29	0.54
18-29 Female	3.57	1.0370	96.21	92.61	85.92	0.45
30-49 Male	3.54	1.0367	96.34	91.85	85.46	0.40
30-49 Female	1.85	1.0188	96.55	93.43	88.57	0.35
50+ Male	0.17	1.0017	96.39	91.32	87.93	0.53
50+ Female	-1.82	0.9821	96.77	92.65	91.26	0.48
Total	1.18	1.0119	97.07	95.28	91.59	0.14
0-17	1.54	1.0157	96.00	95.94	90.84	0.19
18-29 Male	3.77	1.0391	96.35	92.90	86.49	0.34
18-29 Female	2.23	1.0228	96.54	93.62	88.55	0.30
30-49 Male	1.86	1.0190	97.47	95.23	91.24	0.20
30-49 Female	0.96	1.0097	97.63	96.00	92.92	0.17
50+ Male	-0.25	0.9975	97.96	95.35	93.69	0.18
50+ Female	-0.79	0.9922	97.92	95.52	94.30	0.17

Net Undercount is for household population.

CVs for age/sex are approximate.

Table A-5: 2000 A.C.E. Race/Origin by Age/Sex Coverage Estimates

Race/Origin by Age/Sex	Net Undercount (%)	Coverage Correction Factor	Data Defined (%)	Correct Enumeration Rate (%)	Match Rate (%)	CV(CCF) (%)
Non-Hispanic White	0.67	1.0068	97.70	95.90	93.15	0.14
0-17	1.15	1.0117	97.06	96.73	92.85	0.21
18-29 Male	2.38	1.0244	97.03	93.68	88.98	0.37
18-29 Female	1.31	1.0133	97.16	94.14	90.41	0.34
30-49 Male	1.34	1.0136	97.89	96.00	92.79	0.21
30-49 Female	0.79	1.0079	98.02	96.57	93.95	0.18
50+ Male	-0.32	0.9968	98.23	95.74	94.37	0.20
50+ Female	-0.92	0.9909	98.20	95.81	94.97	0.18
Non-Hispanic Black	2.17	1.0221	95.65	92.73	86.93	0.35
0-17	2.99	1.0309	94.67	93.50	85.93	0.50
18-29 Male	4.44	1.0465	95.01	90.18	82.16	0.98
18-29 Female	3.99	1.0416	95.39	91.92	84.21	0.80
30-49 Male	2.85	1.0294	96.10	91.33	85.41	0.63
30-49 Female	1.29	1.0131	96.44	93.46	89.04	0.47
50+ Male	-0.74	0.9927	96.61	92.42	89.98	0.62
50+ Female	-0.87	0.9914	96.80	93.63	91.37	0.56
Hispanic	2.85	1.0294	95.21	94.46	87.48	0.40
0-17	1.77	1.0180	93.66	95.24	87.64	0.51
18-29 Male	7.92	1.0860	95.14	92.30	80.85	0.82
18-29 Female	3.94	1.0411	95.41	93.55	85.75	0.67
30-49 Male	3.84	1.0399	96.35	93.96	87.09	0.55
30-49 Female	1.49	1.0152	96.65	95.40	90.84	0.49
50+ Male	0.47	1.0047	96.59	94.13	90.50	0.69
50+ Female	0.15	1.0015	96.42	94.97	91.45	0.60
Native Hawaiian or PI	4.60	1.0483	95.41	93.05	84.91	2.90
0-17	6.40	1.0684	94.05	93.28	82.12	4.26
18-29 Male	5.08	1.0535	95.02	92.93	84.72	4.66
18-29 Female	3.35	1.0347	95.31	92.71	86.17	4.18
30-49 Male	5.84	1.0620	96.50	93.28	84.41	5.54
30-49 Female	1.33	1.0135	96.75	92.64	87.96	3.46
50+ Male	6.73	1.0721	96.53	93.71	84.05	6.71
50+ Female	-0.73	0.9927	96.31	92.57	89.79	2.83
Non-Hispanic Asian	0.96	1.0097	96.49	94.57	90.43	0.65
0-17	-0.57	0.9943	94.87	95.43	91.04	0.80
18-29 Male	2.22	1.0227	95.69	91.47	85.78	1.48
18-29 Female	1.47	1.0149	96.11	92.08	87.17	1.55
30-49 Male	2.40	1.0246	97.14	94.85	89.94	0.98
30-49 Female	0.78	1.0079	97.50	95.59	92.48	0.77
50+ Male	0.93	1.0094	97.64	94.90	91.81	0.99
50+ Female	0.88	1.0088	97.39	94.99	91.71	0.94
AI On Reservation	4.74	1.0498	94.13	95.81	85.99	1.25
0-17	3.89	1.0405	92.83	96.37	86.01	1.44
18-29 Male	6.91	1.0743	94.10	94.07	82.34	2.61
18-29 Female	6.94	1.0745	94.61	95.16	84.09	2.11
30-49 Male	4.50	1.0471	95.22	94.61	86.06	2.24
30-49 Female	3.89	1.0405	95.55	96.35	88.54	1.77
50+ Male	7.55	1.0816	95.16	96.49	84.89	3.06
50+ Female	3.75	1.0390	95.33	95.76	88.03	2.62
AI Off Reservation	3.28	1.0339	96.24	93.97	87.86	1.38
0-17	4.66	1.0489	95.30	95.43	86.67	2.31
18-29 Male	3.77	1.0392	95.97	90.72	84.67	3.76
18-29 Female	8.52	1.0932	96.19	92.57	81.50	4.40
30-49 Male	0.79	1.0080	96.78	93.27	89.50	2.39
30-49 Female	4.23	1.0441	96.96	93.96	87.25	2.16
50+ Male	-0.42	0.9958	96.93	95.23	92.89	1.97
50+ Female	-1.70	0.9833	96.92	93.73	92.43	1.99
Total	1.18	1.0119	97.07	95.28	91.59	0.14
0-17	1.54	1.0157	96.00	95.94	90.84	0.19
18-29 Male	3.77	1.0391	96.35	92.90	86.49	0.34
18-29 Female	2.23	1.0228	96.54	93.62	88.55	0.30
30-49 Male	1.86	1.0190	97.47	95.22	91.24	0.20
30-49 Female	0.96	1.0097	97.63	96.00	92.92	0.17
50+ Male	-0.25	0.9975	97.96	95.35	93.69	0.18
50+ Female	-0.79	0.9922	97.92	95.52	94.30	0.17

Net Undercount is for household population
CVs for age/sex are approximate

Table A-6 2000 A C E Coverage Estimates by 64 Major Post-Stratum Groups

					Net Undercount (%)	Coverage Correction Factor	Data Defined (%)	Correct Enumeration Rate (%)	Match Rate (%)	CV (%)
Domain 7 (Non-Hispanic White)	Owner	Large MSA	High	N	0 81	1.0081	98 54	97 41	95 23	0 43
				M	0 01	1 0001	98 90	97 43	96 34	0 36
				S	0 36	1.0037	98 74	96 66	95 12	0 87
				W	-0 38	0 9962	98 58	96 17	95 13	0 45
			Low	N	-3 62	0 9650	95 30	93 89	92 74	1 01
				M	-2 61	0 9746	96 03	95 86	94 44	1 40
				S	2 19	1 0223	97 85	96 35	92 24	1 58
				W	1 14	1 0115	97 67	95 40	92 11	2 11
		Medium MSA	High	N	0.30	1 0030	98 84	97 51	96 10	0 86
				M	-0 12	0 9988	98 99	97 92	97 05	0 28
				S	0 46	1 0047	98 64	96 65	94 88	0 42
				W	-0 28	0.9972	98 48	96 98	95.78	0 38
			Low	N	-4.39	0.9579	97 17	95 96	97 37	1.46
				M	-0.33	0.9967	97.85	96 20	94.49	0 84
				S	0 66	1.0066	98 03	95 63	93.14	1 11
				W	1 81	1.0184	97.70	95 01	91.11	2 84
		Small MSA	High	N	-0 25	0.9975	98 79	96.67	95.75	1 33
				M	0.14	1.0014	98 85	97 23	95 99	0 40
				S	0 44	1 0044	98 82	96 93	95 37	0 44
				W	0 30	1 0030	98 53	97.35	95 67	0.57
			Low	N	2.29	1 0234	98 01	97.52	93 38	3 68
				M	2 61	1 0268	97 81	95 52	90 99	2.18
				S	2 09	1 0214	98 11	95 65	92 07	1 10
				W	2 71	1.0279	97 58	95.77	90 97	1 54
		Other	High	N	1.84	1.0187	97 10	97 17	92 72	1 08
				M	-1.11	0 9890	97.60	97.27	95.99	0 39
				S	1 34	1.0136	97 29	96.64	92 78	0.98
				W	0 85	1.0086	95.75	95.15	90 30	1 68
			Low	N	0 56	1 0057	96 53	95 88	92 01	2 18
				M	-0 16	0 9984	96 64	96.21	93 13	1 20
				S	0.15	1 0015	96 20	95 56	91 80	0.66
				W	1.59	1 0161	95 29	94 15	88 43	1 92
	Non-Owner	Large MSA	High		1.82	1 0186	97 06	93 65	89 25	0.64
			Low		1 02	1 0103	94 39	91 65	85 66	1 02
		Medium MSA	High		0.61	1.0061	97 23	92 94	89 89	0 72
			Low		2 83	1.0291	95 96	91.71	85 56	1 28
		Small MSA	High		2 45	1.0251	97 47	94 39	89 83	0 53
			Low		3 61	1 0374	96.37	91.95	85 57	1.29
		Other	High		1 64	1.0166	95.42	93 94	88 31	0.96
			Low		4.08	1.0426	94 19	92.07	83 34	1 74
Domain 4 (Non-Hispanic Black)	Owner	Large/Medium	High		1 63	1 0166	97 13	95 28	91 05	0 57
			Low		-1 31	0.9871	93 73	91 22	86 84	1 23
		Small/Other	High		0 07	1.0007	96 52	94 07	90 79	1 07
			Low		0.46	1 0046	94 47	94 17	88 74	1 86
	Non-Owner	Large/Medium	High		4 18	1.0436	95 75	91 53	84 06	0 69
			Low		3 42	1.0354	92.91	89.76	80 60	1.08
		Small/Other	High		2.64	1 0272	95.97	91.71	85 78	0.99
			Low		0.12	1.0012	94 21	92 27	86 89	2 08
Domain 3 (Hispanic)	Owner	Large/Medium	High		1.46	1.0148	96 24	96 84	91.86	0 53
			Low		0.04	1.0004	93 41	94 23	88.00	1 27
		Small/Other	High		1.66	1 0169	95 28	96 46	90 46	1 02
			Low		1 08	1 0109	93 57	95 44	88.57	2 11
	Non-Owner	Large/Medium	High		3 52	1 0365	95 69	93 78	86 71	0 69
			Low		4 98	1 0524	93 76	90 79	81 10	1.18
		Small/Other	High		4 88	1.0513	94 90	92 30	83 41	1 63
			Low		10.74	1 1203	92 57	92 13	76 16	4.61
Domain 5 (Hawaiian or Pacific Isl)	Owner				2 71	1 0278	95 51	93 79	87.36	3 94
	Non-Owner				6.58	1.0704	95 30	92.33	82.39	4 36
Domain 6 (Non-Hispanic Asian)	Owner				0 55	1 0055	96 87	95.84	92 34	0 87
	Non-Owner				1 58	1.0161	95 90	92 45	87.33	1.00
Domain 1 (Am Ind On Res)	Owner				5 04	1.0531	94 00	95 65	85 43	1.52
	Non-Owner				4 10	1 0428	94 41	96 15	87 08	1 48
Domain 2 (Am. Ind. Off Res)	Owner				1 60	1 0163	96 49	94 55	90 19	1 98
	Non-Owner				5 57	1 0590	95 88	93 16	84 65	2.14

Net Undercount is for household population

Table A-7: Census 2000 A.C.E. 64 Post-Stratum Groups - Percent Net Undercount

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate NE	High Return Rate MW	High Return Rate S	High Return Rate W	Low Return Rate NE	Low Return Rate MW	Low Return Rate S	Low Return Rate W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	0.81	0.01	0.36	-0.38	-3.62	-2.61	2.19	1.14
		Medium MSA MO/MB	0.30	-0.12	0.46	-0.28	-4.39	-0.33	0.66	1.81
		Small MSA & Non-MSA MO/MB	-0.25	0.14	0.44	0.30	2.29	2.61	2.09	2.71
		All Other TEAs	1.84	-1.11	1.34	0.85	0.56	-0.16	0.15	1.59
	Non-Owner	Large MSA MO/MB	1.82				1.02			
		Medium MSA MO/MB	0.61				2.83			
		Small MSA & Non-MSA MO/MB	2.45				3.61			
		All Other TEAs	1.64				4.08			
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB Medium MSA MO/MB	1.63				-1.31			
		Small MSA & Non-MSA MO/MB All Other TEAs	0.07				0.46			
	Non-Owner	Large MSA MO/MB Medium MSA MO/MB	4.18				3.42			
		Small MSA & Non-MSA MO/MB All Other TEAs	2.64				0.12			
Domain 3 (Hispanic)	Owner	Large MSA MO/MB Medium MSA MO/MB	1.46				0.04			
		Small MSA & Non-MSA MO/MB All Other TEAs	1.66				1.08			
	Non-Owner	Large MSA MO/MB Medium MSA MO/MB	3.52				4.98			
		Small MSA & Non-MSA MO/MB All Other TEAs	4.88				10.74			
Domain 5 (Native Hawaiian or Pacific Islander)	Owner		2.71							
	Non-Owner		6.58							
Domain 6 (Non-Hispanic Asian)	Owner		0.55							
	Non-Owner		1.58							
American Indian or Alaska Native — **Domain 1** (On Reservation)	Owner		5.04							
	Non-Owner		4.10							
American Indian or Alaska Native — **Domain 2** (Off Reservation)	Owner		1.60							
	Non-Owner		5.57							

- For Census 2000, persons can self-identify with more than one race group. For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain. This classification does not change a person's actual response. Further, all official tabulations are based on actual responses to the census.

Table A-8: Census 2000 A.C.E. 64 Post-Stratum Groups - Percent CV

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate NE	High Return Rate MW	High Return Rate S	High Return Rate W	Low Return Rate NE	Low Return Rate MW	Low Return Rate S	Low Return Rate W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	0.43	0.36	0.87	0.45	1.01	1.40	1.58	2.11
		Medium MSA MO/MB	0.86	0.28	0.42	0.38	1.46	0.84	1.11	2.84
		Small MSA & Non-MSA MO/MB	1.33	0.40	0.44	0.57	3.68	2.18	1.10	1.54
		All Other TEAs	1.08	0.39	0.98	1.68	2.18	1.20	0.66	1.92
	Non-Owner	Large MSA MO/MB	0.64				1.02			
		Medium MSA MO/MB	0.72				1.28			
		Small MSA & Non-MSA MO/MB	0.53				1.29			
		All Other TEAs	0.96				1.74			
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB Medium MSA MO/MB	0.57				1.23			
		Small MSA & Non-MSA MO/MB All Other TEAs	1.07				1.86			
	Non-Owner	Large MSA MO/MB Medium MSA MO/MB	0.69				1.08			
		Small MSA & Non-MSA MO/MB All Other TEAs	0.99				2.08			
Domain 3 (Hispanic)	Owner	Large MSA MO/MB Medium MSA MO/MB	0.53				1.27			
		Small MSA & Non-MSA MO/MB All Other TEAs	1.02				2.11			
	Non-Owner	Large MSA MO/MB Medium MSA MO/MB	0.69				1.18			
		Small MSA & Non-MSA MO/MB All Other TEAs	1.63				4.61			
Domain 5 (Native Hawaiian or Pacific Islander)	Owner		3.94							
	Non-Owner		4.36							
Domain 6 (Non-Hispanic Asian)	Owner		0.87							
	Non-Owner		1.00							
American Indian or Alaska Native – **Domain 1** (On Reservation)	Owner		1.52							
	Non-Owner		1.48							
American Indian or Alaska Native – **Domain 2** (Off Reservation)	Owner		1.98							
	Non-Owner		2.14							

* For Census 2000, persons can self-identify with more than one race group. For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain. This classification does not change a person's actual response. Further, all official tabulations are based on actual responses to the census

Table A-9: Census 2000 A.C.E. 64 Post-Stratum Groups -Census Counts (Millions)

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate NE	High Return Rate MW	High Return Rate S	High Return Rate W	Low Return Rate NE	Low Return Rate MW	Low Return Rate S	Low Return Rate W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	11 5	6 8	5 3	5 8	3.0	0 7	1 5	0 9
		Medium MSA MO/MB	5.6	11.7	11 9	8 2	0.5	1 1	3 6	2 3
		Small MSA & Non-MSA MO/MB	2.9	11 4	7 9	3 6	0 4	0.9	3 4	1 5
		All Other TEAs	4 1	6 9	5 0	1.8	1 4	1.0	11 9	2 3
	Non-Owner	Large MSA MO/MB	8 5				3 8			
		Medium MSA MO/MB	11 5				3.1			
		Small MSA & Non-MSA MO/MB	9.4				2 1			
		All Other TEAs	5.0				2.5			
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB	8 5				2.8			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	3 9				1.3			
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	9 2				3.7			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	3.5				0.5			
		All Other TEAs								
Domain 3 (Hispanic)	Owner	Large MSA MO/MB	9 7				2.5			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	2.9				1.7			
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	10 7				3 8			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	2 6				0.7			
		All Other TEAs								
Domain 5 (Native Hawaiian or Pacific Islander)	Owner		0.3							
	Non-Owner		0.3							
Domain 6 (Non-Hispanic Asian)	Owner		6 0							
	Non-Owner		3.9							
American Indian or Alaska Native — **Domain 1** (On Reservation)	Owner		0 4							
	Non-Owner		0 2							
American Indian or Alaska Native — **Domain 2** (Off Reservation)	Owner		0.9							
	Non-Owner		0 6							

- For Census 2000, persons can self-identify with more than one race group For post-stratification purposes, persons are included in a single Race/Hispanic Origin Domain. This classification does not change a person's actual response Further, all official tabulations are based on actual responses to the census.

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Some Bookkeeping Details

- The final 2000 A.C.E. post-stratum design has 416 direct estimates.
- For 8 post-stratum groups, 7 age/sex cells were collapsed into 3 age/sex cells.
    - 7 occurrences due to small sample sizes (< 100 P-sample persons)
    - 1 occurrence due to an outlier CV
- PES-A implemented 63 times out of 416 post-strata (< 10 outmovers)

Race/Hispanic Origin Domains

The Race/Hispanic origin domain assignment is hierarchical. See Haines (2000) for more detail.

Domain 1 (American Indian or Alaska Native on reservations) includes:

- Any person living on a reservation indicating American Indian or Alaska Native either as their single race or as one of many races, regardless of their Hispanic origin.

Domain 2 (American Indian or Alaska Native off reservations) includes:

- Any person living in Indian Country[1] but not on a reservation who indicates American Indian or Alaska Native either as their single race or as one of many races, regardless of their Hispanic origin.
- Any non-Hispanic person not living in Indian Country who indicates American Indian or Alaska Native as their single race.

Domain 3 (Hispanic) includes:

- All Hispanic persons who are not included in Domains 1 or 2.
- All Hispanic persons who self-identify with three or more races (excluding American Indian or Alaska Native in Indian Country).
- All Hispanic persons who do not live in the state of Hawaii who classify themselves as Native Hawaiian or Pacific Islander, regardless of whether they identify with a single or multiple race.

[1] Indian Country is land considered (either wholly or partially) on an American Indian reservation/trust land, Tribal Jurisdiction Statistical Area, Tribal Designated Statistical Area, or Alaska Native Village Statistical Area. For Census 2000, Tribal Jurisdiction Statistical Area has been formally renamed as Oklahoma Tribal Statistical Area.

Domain 4 (Non-Hispanic Black) includes:

- Any non-Hispanic person who indicates Black as their only race.
- Any person identifying with a combination of Black and American Indian or Alaska Native not in Indian Country.
- Any person who indicates Black and another single race group (Native Hawaiian or Pacific Islander, Asian, White, or "Some other race").
- All Non-Hispanic Black persons who do not live in the state of Hawaii who classify themselves as Native Hawaiian or Pacific Islander.

Domain 5 (Native Hawaiian or Pacific Islander) includes:

- Any non-Hispanic person indicating the single race Native Hawaiian or Pacific Islander.
- Any non-Hispanic person who identifies with the race combination of Native Hawaiian or Pacific Islander and American Indian or Alaska Native not in Indian Country.
- Any non-Hispanic person who identifies with the race combination of Native Hawaiian or Pacific Islander and Asian.
- All persons living in the state of Hawaii who classify themselves as Native Hawaiian or Pacific Islander, regardless of their Hispanic origin and whether they identify with a single or multiple race.

Domain 6 (Non-Hispanic Asian) includes:

- Any non-Hispanic person indicating Asian as their single race.
- Any person who self-identifies with Asian and American Indian or Alaska Native not in Indian Country.

Domain 7 (Non-Hispanic White or "Some other race") includes:

- Any Non-Hispanic person indicating White or "Some other race" as their single race.
- Any Non-Hispanic person who self-identifies with both American Indian or Alaska Native not in Indian Country and White or "Some other race."
- Any person who self-identifies with Asian and White or Asian and "Some other race."
- Any non-Hispanic person who self-identifies with three or more races (excluding American Indian or Alaska Native in Indian Country).
- Any Non-Hispanic White or Non-Hispanic "Some other race" person who classifies themselves as Native Hawaiian or Pacific Islander but does not live in Hawaii, regardless of whether they identify with other races.

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Table P-1: 1990 PES Coverage Estimates

Characteristic	Net Undercount[1] (%)	Coverage Correction Factor	Data Defined (%)	Correct Enumeration Rate (%)	Match Rate (%)	CV(CCF) (%)
Total	1.61	1.0163	99.10	94.26	92.22	0.20
Race/Origin						
White/Other	0.68	1.0068	99.27	94.97	93.80	0.22
Black	4.57	1.0479	98.43	90.90	85.81	0.55
Hispanic	4.99	1.0526	98.68	92.93	87.37	0.82
API	2.36	1.0241	98.84	92.78	90.18	1.39
Am. Ind. on Res.	12.22	1.1392	96.84	91.54	78.13	5.29
Tenure[2]						
Owner	0.04	1.0004	99.44	95.51	94.95	0.21
Non-Owner	4.51	1.0472	98.45	91.69	86.41	0.43
Age/Sex						
0 - 17	3.18	1.0329	99.18	95.08	91.54	0.29
18 - 29 Male	3.30	1.0341	98.55	90.33	86.70	0.54
18 - 29 Female	2.83	1.0291	98.55	91.92	88.36	0.47
30 - 49 Male	1.89	1.0193	99.24	94.14	92.14	0.32
30 - 49 Female	0.88	1.0089	99.30	95.20	93.81	0.25
50+ Male	-0.59	0.9942	99.31	95.17	95.18	0.34
50+ Female	-1.24	0.9878	99.21	95.41	95.89	0.29

[1] Net undercount is for PES universe.

[2] Excludes American Indians on Reservations.

Table P-2: 1990 PES Net Undercount Estimate Distribution

Characteristic	Net Undercount[1] (x 1000)	Distribution (%)
Total	3994	100
Race/Origin		
Non-Hispanic White/Other	1277	32
Black	1389	35
Non-Black Hispanic	1102	28
Asian and Pac. Islander	174	4
AI On Reservation	52	1
Tenure[2]		
Owner	71	2
Non-Owner	3871	97
Age/Sex		
0 - 17	2084	52
18 - 29 Male	792	20
18 - 29 Female	687	17
30 - 49 Male	685	17
30 - 49 Female	326	8
50+ Male	-160	-4
50+ Female	-419	-10

[1] Net undercount is for PES universe.

[2] Excludes American Indians on Reservations.

Table P-3: 1990 PES Race/Origin by Tenure Coverage Estimates

Race/Origin by Tenure	Net Undercount[1] (%)	Coverage Correction Factor	Data Defined (%)	Correct Enumeration Rate (%)	Match Rate (%)	CV(CCF) (%)
White/Other	0.68	1.0068	99.27	94.97	93.80	0.22
Owner	-0.26	0.9974	99.54	95.84	95.64	0.23
Non-Owner	3.06	1.0315	98.56	92.61	88.62	0.51
Black	4.57	1.0479	98.43	90.90	85.81	0.55
Owner	2.25	1.0231	98.80	92.84	89.65	0.58
Non-Owner	6.48	1.0693	98.11	89.19	82.28	0.88
Hispanic	4.99	1.0526	98.68	92.93	87.37	0.82
Owner	1.82	1.0185	98.97	95.56	92.81	0.69
Non-Owner	7.43	1.0803	98.44	90.58	82.45	1.28
API	2.36	1.0241	98.84	92.78	90.18	1.39
Owner	-1.45	0.9857	99.26	93.13	93.71	1.49
Non-Owner	6.96	1.0748	98.29	92.22	84.36	2.68
Total [2]	1.59	1.0162	99.10	94.27	92.24	0.20
Owner	0.04	1.0004	99.44	95.51	94.95	0.21
Non-Owner	4.51	1.0472	98.45	91.69	86.41	0.43

[1] Net undercount is for PES universe.

[2] Excludes American Indians on Reservations.

Table P-4: 1990 PES Tenure by Age/Sex Coverage Estimates

Tenure by Age/Sex	Net Undercount[1] (%)	Coverage Correction Factor	Data Defined (%)	Correct Enumeration Rate (%)	Match Rate (%)	CV(CCF) (%)
Owners[2]	0.04	1.0004	99.44	95.51	94.95	0.21
0 - 17	1.56	1.0159	99.37	96.34	94.31	0.32
18 - 29 Male	-0.46	0.9954	99.28	91.24	90.98	0.56
18 - 29 Female	-0.38	0.9962	99.29	91.89	91.67	0.57
30 - 49 Male	-0.08	0.9992	99.50	95.50	95.16	0.31
30 - 49 Female	0.14	1.0014	99.53	96.49	95.90	0.23
50+ Male	-1.06	0.9895	99.53	96.12	96.67	0.30
50+ Female	-0.96	0.9905	99.45	96.27	96.69	0.31
Non-Owners[2]	4.51	1.0472	98.45	91.69	86.41	0.43
0 - 17	5.83	1.0619	98.85	92.75	86.42	0.60
18 - 29 Male	6.61	1.0708	97.86	89.32	81.80	0.87
18 - 29 Female	5.46	1.0578	97.90	91.94	84.99	0.74
30 - 49 Male	5.94	1.0631	98.66	91.00	84.92	0.76
30 - 49 Female	2.48	1.0254	98.79	92.17	88.80	0.64
50+ Male	1.52	1.0154	98.28	90.82	88.10	1.24
50+ Female	-2.27	0.9778	98.34	92.48	92.96	0.73
Total	1.61	1.0163	99.10	94.26	92.22	0.20
0 - 17	3.18	1.0329	99.18	95.08	91.54	0.29
18 - 29 Male	3.30	1.0341	98.55	90.33	86.70	0.54
18 - 29 Female	2.83	1.0291	98.55	91.92	88.36	0.47
30 - 49 Male	1.89	1.0193	99.24	94.14	92.14	0.32
30 - 49 Female	0.88	1.0089	99.30	95.20	93.81	0.25
50+ Male	-0.59	0.9942	99.31	95.17	95.18	0.34
50+ Female	-1.24	0.9878	99.21	95.41	95.89	0.29

[1] Net undercount is for PES universe.

[1] Excludes American Indians on Reservations.

Table P-5: 1990 PES Race/Origin by Age/Sex Coverage Estimates

Race/Origin by Age/Sex		Net Undercount [1] (%)	Coverage Correction Factor	Data Defined (%)	Correct Enumeration Rate (%)	Match Rate (%)	CV(CCF) (%)
White/Other		0.68	1.0068	99.27	94.97	93.80	0.22
0 - 17		1.96	1.0200	99.45	95.97	93.72	0.33
18 - 29	Male	2.12	1.0216	98.68	91.17	88.45	0.61
18 - 29	Female	1.88	1.0192	98.67	92.54	89.90	0.54
30 - 49	Male	0.83	1.0084	99.38	94.81	93.70	0.36
30 - 49	Female	0.17	1.0017	99.44	95.81	95.15	0.28
50+	Male	-0.83	0.9918	99.40	95.64	95.90	0.35
50+	Female	-1.38	0.9864	99.31	95.84	96.52	0.30
Black		4.57	1.0479	98.43	90.90	85.81	0.55
0 - 17		7.09	1.0763	98.53	92.03	84.37	0.94
18 - 29	Male	4.04	1.0421	98.05	84.95	80.89	1.35
18 - 29	Female	5.57	1.0589	98.05	89.62	83.29	1.20
30 - 49	Male	6.75	1.0724	98.48	89.80	83.12	0.99
30 - 49	Female	3.23	1.0333	98.60	92.41	88.28	0.68
50+	Male	-0.39	0.9961	98.55	91.39	90.30	0.97
50+	Female	-1.27	0.9875	98.46	93.08	92.86	0.61
Hispanic		4.99	1.0526	98.68	92.93	87.37	0.82
0 - 17		4.89	1.0515	98.54	93.73	87.88	1.00
18 - 29	Male	7.18	1.0774	98.42	89.63	81.95	2.33
18 - 29	Female	5.76	1.0611	98.52	92.27	85.31	1.25
30 - 49	Male	6.28	1.0670	98.87	92.87	86.43	1.35
30 - 49	Female	3.69	1.0383	98.99	93.14	88.97	1.16
50+	Male	2.73	1.0280	98.96	93.31	91.18	2.11
50+	Female	2.18	1.0223	98.90	94.95	92.24	1.98
API		2.36	1.0241	98.84	92.78	90.18	1.39
0 - 17		3.26	1.0337	98.85	93.73	90.29	2.17
18 - 29	Male	10.47	1.1170	98.06	92.93	83.25	4.09
18 - 29	Female	1.71	1.0174	98.21	86.95	84.41	3.99
30 - 49	Male	-0.10	0.9990	99.09	92.90	92.86	2.14
30 - 49	Female	1.88	1.0191	99.22	95.09	93.17	1.50
50+	Male	0.04	1.0004	99.17	95.38	94.61	1.29
50+	Female	-3.74	0.9639	99.11	89.09	91.72	2.82
American Indian on Reservations		12.22	1.1392	96.84	91.54	78.13	5.29
0 - 17		13.76	1.1595	96.59	92.15	76.77	5.72
18 - 29	Male	17.29	1.2091	96.39	89.19	71.11	8.05
18 - 29	Female	15.27	1.1802	96.67	92.21	75.53	7.28
30 - 49	Male	14.08	1.1639	97.07	92.31	76.99	4.79
30 - 49	Female	4.30	1.0449	97.34	90.45	84.27	5.85
50+	Male	8.80	1.0965	97.47	91.93	81.73	5.75
50+	Female	3.90	1.0405	97.28	90.94	85.02	4.03
Total		1.61	1.0163	99.10	94.26	92.22	0.20
0 - 17		3.18	1.0329	99.18	95.08	91.54	0.29
18 - 29	Male	3.30	1.0341	98.55	90.33	86.70	0.54
18 - 29	Female	2.83	1.0291	98.55	91.92	88.36	0.47
30 - 49	Male	1.89	1.0193	99.24	94.14	92.14	0.32
30 - 49	Female	0.88	1.0089	99.30	95.20	93.81	0.25
50+	Male	-0.59	0.9942	99.31	95.17	95.18	0.34
50+	Female	-1.24	0.9878	99.21	95.41	95.89	0.29

[1] Net undercount is for PES universe.

Table P-6: 1990 PES Coverage Estimates by 51 Major Post-Stratum Groups

				Net Undercount (%)	Coverage Correction Factor	Data Defined (%)	Correct Enumeration Rate (%)	Match Rate (%)	CV (%)
White/Other	Owner	Large City	Northeast	-2.13	0.9792	99.48	94.96	96.44	1.06
			Midwest	-0.26	0.9974	99.73	97.19	97.15	0.40
			South	0.68	1.0069	99.50	96.06	94.84	0.71
			West	-0.34	0.9966	99.53	96.57	96.42	0.65
		Medium City	Northeast	-1.08	0.9893	99.55	96.92	97.45	0.48
			Midwest	-0.10	0.9990	99.73	97.40	97.14	0.40
			South	0.52	1.0052	99.57	96.22	95.25	0.43
			West	0.62	1.0063	99.44	96.11	95.00	0.58
		Rural	Northeast	-0.54	0.9947	99.49	95.09	95.15	0.64
			Midwest	-0.71	0.9930	99.63	95.01	95.38	1.18
			South	0.18	1.0018	99.43	93.96	93.30	0.70
			West	0.29	1.0029	99.08	95.26	94.17	0.68
	Non-Owner	Large City	Northeast	1.16	1.0117	98.17	90.44	87.67	1.40
			Midwest	2.33	1.0239	98.69	94.24	91.08	1.66
			South	2.56	1.0262	98.39	90.94	87.31	1.51
			West	3.18	1.0328	98.74	92.79	89.02	1.66
		Medium City	Northeast	3.41	1.0353	98.03	94.49	89.42	1.50
			Midwest	1.23	1.0124	98.28	94.56	91.92	1.12
			South	3.20	1.0330	98.51	92.20	88.02	1.80
			West	4.49	1.0470	98.70	93.77	88.46	1.40
		Rural	Northeast	6.52	1.0697	98.84	91.95	85.01	4.40
			Midwest	2.85	1.0293	99.32	92.67	89.63	1.57
			South	6.23	1.0665	99.04	92.87	86.27	1.81
			West	6.08	1.0648	98.67	94.12	87.36	2.04
Black	Owner	Large City	Northeast	1.63	1.0165	98.15	89.15	86.22	1.94
			Midwest	0.81	1.0082	98.83	93.25	91.52	0.87
			South	2.16	1.0221	98.83	94.35	91.38	0.92
			West	6.10	1.0649	98.92	91.69	85.27	2.02
		Medium City		1.34	1.0136	99.06	93.26	91.42	1.00
		Rural		3.52	1.0364	98.87	92.73	88.31	1.97
	Non-Owner	Large City	Northeast	8.37	1.0913	97.50	84.69	76.18	1.76
			Midwest	5.99	1.0638	98.06	89.68	82.77	1.79
			South	6.27	1.0669	98.16	89.05	82.01	2.04
			West	9.96	1.1106	98.41	89.79	80.67	3.02
		Medium City		4.15	1.0433	98.44	92.24	87.27	1.23
		Rural		4.62	1.0484	98.47	88.60	83.47	5.71
Hispanic	Owner	Large City	Northeast	0.67	1.0068	98.68	93.11	91.09	4.39
			Midwest	-4.33	0.9585	99.15	92.51	95.68	2.48
			South	2.53	1.0259	99.22	95.69	92.74	0.92
			West	2.89	1.0298	98.93	97.00	93.34	0.91
		Medium City		0.94	1.0095	98.97	95.79	93.99	1.68
		Rural		2.73	1.0280	98.68	92.29	89.04	2.80
	Non-Owner	Large City	Northeast	6.72	1.0721	98.37	84.92	78.26	3.76
			Midwest	6.64	1.0711	98.17	88.00	80.62	3.49
			South	9.34	1.1030	98.73	93.67	83.97	2.80
			West	5.91	1.0629	98.42	91.17	84.53	1.98
		Medium City		6.60	1.0707	98.45	91.98	84.63	2.92
		Rural		15.80	1.1876	98.08	89.26	73.98	6.11
API	Owner			-1.45	0.9857	99.26	93.13	93.71	1.49
	Non-Owner			6.96	1.0748	98.29	92.22	84.36	2.68
AI on Reservations				12.22	1.1392	96.84	91.54	78.13	5.29

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Decomposition of the Dual System Estimate Components

Why decompose the components of the Dual System Estimates (DSEs)?

- We decompose to show the outcome of the estimation steps on the DSEs by documenting how these steps contribute to the estimated components of dual system estimation.

- This provides another source of verification for estimates of the components.

What estimated components will be examined?

We examine the following estimated components:

E sample

- Correct Enumerations
- Enumerations

P sample

- Nonmovers
- Nonmover Matches

- Outmovers
- Outmover Matches

- Inmovers

What estimation steps are examined?

We examined four steps:

1. Noninterview adjustments for whole-household nonresponse in the P sample

   There are two noninterview adjustments:

   - Noninterview adjustment for Census Day Interview: This is used to adjust the weights of nonmovers and outmovers.

   - Noninterview adjustment for A.C.E. Interview Day: This is used to adjust the weights of inmovers.

   Note: This step only applies to the P sample.

2. Characteristic imputation of missing post-stratification variables for the P and E samples

   - Variables requiring imputation for post-stratification for both the P and E samples

     - Age
     - Hispanic Origin
     - Race
     - Sex
     - Tenure

3. Imputation of Status (Resident, Match and Enumeration)

   - This assigns probabilities to people with unresolved resident and match status for the P sample.

   - This assigns probabilities to people with unresolved enumeration status for the E sample.

   - Imputation Cell Estimation (ICE) determines the probabilities.

4. Targeted Extended Search (TES) which expanded the search area for selected cases during person matching to reduce the variance of the DSEs.

   - This steps applies to both the P and E samples.

How are the estimation steps grouped for this analysis?

We formed 11 groups:

Table 1: Groupings for Decomposition Analysis

Number	Group	P Sample	E Sample
1	Resolved Cases: Prior to Missing Data Adjustments (Non-TES)	✓	✓
2	Resolved Cases: Added by A.C.E. Interview Day Non-interview Adjustment (Non-TES)	✓	
3	Resolved Cases: Added by Census Day Non-interview Adjustment (Non-TES)	✓	
4	Characteristic Imputation Only (Non-TES)	✓	✓
5	Characteristic Imputation and Residence ICE (Non-TES)	✓	
6	Residence ICE Only (Non-TES)	✓	
7	Characteristic Imputation, Residence ICE and Match ICE (Non-TES)	✓	
8	Residence ICE and Match ICE (Non-TES)	✓	
9	Characteristic Imputation and Correct Enumeration ICE (Non-TES)		✓
10	Correct Enumeration ICE Only (Non-TES)		✓
11	TES	✓	✓

National Summary Table

				E Sample				P Sample									
				Correct Enumerations		Enumerations		Nonmover Matches		Outmover Matches		Nonmovers		Inmovers		Outmovers	
Non-TES	1	Resolved Cases Prior to Missing Adj	Sample Size	555,600		571,503		517,810		12,613		550,430		35,232		15,335	
			Unweighted	555,307 6		571,503.0		517,810.0		12,613 0		550,430 0		35,232.0		15,335.0	
			Weighted	212,578,646 1	84 3%	218,194,828 5	82.5%	204,457,554 7	88 7%	4,796,251 1	70 0%	215,855,256 7	86 4%	12,671,965.1	95 0%	5,797,399.5	65 6%
	2	Resolved Cases Added A C E Day NI	Weighted	0 0	0 0%	0 0	0 0%	0 0	0 0%	0 0	0 0%	0 0	0 0%	157,634 5	1 2%	0 0	0 0%
	3	Resolved Cases Added Census Day NI	Weighted	0 0	0 0%	0 0	0 0%	5,243,017 1	2 3%	152,479 7	2 2%	5,647,569 0	2 3%	0 0	0 0%	184,497 2	2 1%
	4	Characteristic Imputation Only	Sample Size	78,044		91,034		23,568		1,870		26,160		1,391		2,243	
			Unweighted	77,978.0		91,034.0		23,568 0		1,870 0		26,160.0		1,391 0		2,243.0	
			Weighted	27,462,501 1	10 9%	31,982,497 6	12 1%	9,454,845 2	4.1%	735,227 3	10 7%	10,416,980 4	4 2%	502,475 5	3 8%	870,401 9	9 8%
	5	Characteristic Imputation and Residence ICE	Sample Size	0		0		200		218		802		0		468	
			Unweighted	0.0		0 0		176.1		203 4		604 1		0 0		370 5	
			Weighted	0 0	0 0%	0 0	0 0%	75,791 7	0 0%	81,620 3	1.2%	237,256 4	0.1%	0 0	0 0%	150,320 0	1 7%
	6	Residence ICE Only	Sample Size	0		0		266		270		3,975		0		1,292	
			Unweighted	0.0		0 0		242 5		240.0		2,632 5		0 0		866 0	
			Weighted	0 0	0 0%	0 0	0 0%	100,704 3	0 0%	92,713 1	1 4%	968,624 2	0 4%	0 0	0 0%	330,788 9	3 7%
	7	Charactistic Imputation, Residence ICE and Match ICE	Sample Size	0		0		2,835		965		2,835		0		965	
			Unweighted	0.0		0 0		2,031 5		609 4		2,385.7		0 0		820 0	
			Weighted	0 0	0 0%	0 0	0 0%	821,702 6	0 4%	230,500 2	3 4%	956,074 4	0 4%	0 0	0 0%	306,963 5	3 5%
	8	Residence ICE and Match ICE	Sample Size	0		0		2,519		1,499		2,519		0		1,499	
			Unweighted	0.0		0 0		1,941 5		949.0		2,144 7		0 0		1,278 7	
			Weighted	0 0	0 0%	0 0	0 0%	789,961 9	0 3%	366,771 8	5 4%	866,661 6	0 3%	0 0	0 0%	488,806 0	5 5%
	9	Characteristic Imputation and Correct Enumeration ICE	Sample Size	3,353		3,356		0		0		0		0		0	
			Unweighted	2,437.3		3,356 0		0 0		0 0		0 0		0 0		0 0	
			Weighted	754,883 8	0 3%	1,045,557 8	0 4%	0 0	0 0%	0 0	0 0%	0 0	0 0%	0 0	0 0%	0 0	0 0%
	10	Correct Enumeration ICE only	Sample Size	14,661		14,673		0		0		0		0		0	
			Unweighted	11,177.0		14,673 0		0 0		0 0		0 0		0.0		0 0	
			Weighted	3,415,672 8	1 4%	4,625,803 6	1 7%	0 0	0 0%	0 0	0 0%	0 0	0 0%	0 0	0 0%	0 0	0 0%
TES	11	TES Cases	Sample Size	21,870		32,334		20,947		941		45,193		0		2,356	
			Unweighted	21,416 9		32,334 0		20,939 2		938 0		45,009 8		0 0		2,276 2	
			Weighted	7,884,534 5	3 1%	8,730,175 0	3 3%	9,609,355 0	4 2%	392,717 7	5 7%	14,756,760 9	5 9%	0.0	0 0%	713,021 4	8 1%
Total	12	Total	Sample Size	673,528		712,900		568,145		18,376		631,914		36,623		24,158	
			Unweighted	668,316.6		712,900 0		566,708 7		17,422 8		629,366 7		36,623.0		23,189 4	
			Weighted	252,096,238 3	100 0%	264,578,862 5	100 0%	230,552,932 4	100 0%	6,848,281 2	100 0%	249,705,183 6	100 0%	13,332,075 2	100 0%	8,842,198 4	100.0%

ESCAP MEETING NO. 30 - 01/26/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 30
January 26, 2001

Prepared by: Nick Birnbaum

The thirtieth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on January 26, 2001 at 10:30.

The agenda for the meeting was to discuss the preliminary Dual System Estimation (DSE) results and their standard errors, and the decomposition of the DSE components.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines	Tom Mule
Donna Kostanich	Roxie Jones
Raj Singh	Kathleen Styles
David Whitford	Nick Birnbaum
Deborah Fenstermaker	Sarah Brady
Gregg Robinson	Carolee Bush
Danny Childers	Annette Quinlan
Dawn Haines	Maria Urrutia

I. Preliminary DSE Results and Their Reliability

Dawn Haines presented the preliminary DSE results to the Committee. The results were discussed, compared to 1990 PES results, and a number of observations were made:

- As measured by the A.C.E., the differential undercount for Blacks and Hispanics has been reduced approximately by half.

- The standard errors of the coverage correction factors (CCFs) for almost all race/origin, tenure, and age/sex groupings at the national level are low and do not show any unexplained anomalies. The standard errors are somewhat larger for Native Americans on and off reservations and Native Hawaiians.

- Considerable gains have been made in reducing undercoverage from the last census. This result holds for children and non-Hispanic Blacks and Hispanic of any race.

- Tenure continues to follow historical patterns and is a powerful variable in predicting patterns of undercount.

- Non-Hispanic White owners in mailout/mailback portions of large and medium MSAs in the Northeast and Midwest where the mail return rate was low, had high estimated net overcount rates compared to their counterparts with high mail return rates, who had small estimated net overcount or undercount rates.

An important next step for assessing the A.C.E. estimates is to examine the loss functions to see if there is improvement (vis-a-vis the initial census numbers) at the congressional district level.

The Committee examined the percent of data defined persons for different population groups and noted that this percentage had fallen in total and across race/origin, tenure, and age/sex groupings from 1990. This result was due to an increase of persons in substituted households for 2000. The Committee also examined data on correct enumeration and match rates and noted that these data were similar to the corresponding data from 1990. However, the duplication rate was lower than for 1990. Persons in housing units identified as potential duplicates were not included in the A.C.E. matching, therefore, the effect on the A.C.E. estimates of the duplicate processing -- including the reinstatement of a portion of the housing units originally identified as duplicate -- will need to be further researched by DSSD staff.

There was a brief discussion of the collapsing of the 448 post-strata down to 416 post-strata. For eight post-stratum groups, the seven age/sex groupings were collapsed into three groupings. For seven of these post-stratum groups, the collapsing was done because of small sample sizes, and for the eighth post-stratum group, the collapsing was implemented as a result of large variance in the post-stratum. The collapsing based on variance, while not pre-specified, was determined to be appropriate and beneficial. (See discussion in Section I of the January 10, 2001, ESCAP meeting minutes.)

The pre-specified procedures for movers stipulated that in situations where there aren't enough outmovers to support the full mover treatment a revised methodology would be used. A complete explanation of these procedures is found in the DSSD Census 2000 Procedures and Operations Memorandum Series Number Q-37. This revised procedure was used in about 15 percent of the 416 post-strata.

The Committee also noted that it will be important to explain the differences between the A.C.E. results and the demographic analysis (DA) estimates (preliminary DA estimates were presented to the Committee in December).

II. Decomposition of DSE Components

Tom Mule explained, to the Committee, the decomposition of the DSE components (see Attachment 3). The DSE components are decomposed to illustrate the outcome of the estimation steps on the dual system estimates by documenting how the steps contributed to the estimated components of dual system estimation. Decomposition also provides a means for verifying the estimated components of the DSEs.

The Committee briefly examined and discussed the results of the decomposition analysis. It was noted that an examination of the Targeted Extended Search (TES) results (an important component of the decomposition analysis) was already scheduled as an agenda item for a future meeting.

III. Next Meeting

The agenda for the next meeting, to be held on February 1, 2001, is to discuss census quality measures.

ESCAP MEETING NO. 31 - 02/01/01

AGENDA

Kathleen P Porter
01/31/2001 03:05 PM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Jeannette D Greene/POL/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc: James B Treat/DSSD/HQ/BOC@BOC, Jennifer W Reichert/DSSD/HQ/BOC@BOC, Nicholas S Alberti/DSSD/HQ/BOC@BOC

Subject: Agenda for 2/1 ESCAP

The agenda for the February 1 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

Census Quality

1. Address List Development - Jim Treat and Jennifer Reichert

2. Respondent Cooperation - Jim Treat

3. Followup Operations - Nick Alberti and Jennifer Reichert

4. Processing - Jim Treat and Nick Alberti

5. Completeness of the Data (Housing Units Only) - Jim Treat and Nick Alberti

ESCAP MEETING NO. 31 - 02/01/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Preliminary Results from Census 2000

Quality Indicators

for the

Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP)

February 1, 2001

I. Address List Development

A. Descriptive Statistics

In this section we give data on housing unit (HU) records that have been delivered to the Decennial Master Address File (DMAF), the address list for Census 2000. The HU records are classified according to when they were added to the DMAF - either Pre- Questionnaire Delivery, at the time of Questionnaire Delivery, or Post-Questionnaire Delivery.

This classification of HU records by relationship to Questionnaire Delivery operations was based on which address list-building operation was initially responsible for adding the address to the list, known as the Original Source of an address. For the address list-building operations, the country was divided into Type of Enumeration Areas (TEAs), depending on address types and the need for special enumeration or questionnaire delivery procedures. Not every address list-building operation occurred in every TEA. However within a TEA there was still overlap in the timing of the address list-building operations. When an address was independently added by two or more overlapping operations, the Original Source is a combination of those address list-building operations. The Original Source is determined according to the specifications in the Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: MAF-EXT-S-01, subject: "Determining Original Source for the November 2000 Master Address File for Evaluation Purposes" (draft).

Possible operations or files that can add an address to the Census address list are:

The 1990 Address Control File (ACF) (1990 Census address list)
The Delivery Sequence Files (DSFs) from the United States Postal Service
The Census 2000 Dress Rehearsal
Local Update of Census Addresses (LUCA) 1998
(and Supplemental LUCA 1998)
Block Canvassing
Address Listing
LUCA 1999 Relisting
LUCA 1999 Appeals
New Construction
Questionnaire Delivery - in Update/Leave, Urban Update/Leave,
Update/Enumerate, List/Enumerate, and Remote Alaska areas
Special Place/Group Quarters Enumeration
Nonresponse Followup (NRFU)
Coverage Improvement Followup (CIFU)
Be Counted Program

Telephone Questionnaire Assistance Program

The DSFs are address files updated by the United States Postal Service every month. The Census Bureau has incorporated a number of these into the Master Address File, the Census Bureau's address file. These are designated in the Original Source memo according to date delivered: November 1997 (DSF 1), September 1998 (DSF 2), November 1999 (DSF 3), February 2000 (DSF 4), and April 2000 (DSF 5).

The Pre-Questionnaire Delivery category consists of the Original Source operations for which the generated addresses were delivered to the DMAF in advance of Census Day. These are:

1990 ACF
Dress Rehearsal
LUCA 1998
Block Canvassing
Block Canvassing and LUCA 1998
Block Canvassing (Supplemental LUCA)
DSF 1
DSF 2
DSF 3
LUCA 1998 and DSF 2
Address Listing
LUCA 1999 Relisting
LUCA 1999 Appeals
Special Place/Group Quarters Enumeration

The Questionnaire Delivery category consists of all the Original Source operations that include the questionnaire delivery. These are:

Questionnaire Delivery
DSF 5, New Construction and Questionnaire Delivery
LUCA 1999 Appeals and Questionnaire Delivery
New Construction and Questionnaire Delivery
DSF 5 and Questionnaire Delivery
New Construction, NRFU and Questionnaire Delivery
DSF 5, New Construction, NRFU, and Questionnaire Delivery

The Post-Questionnaire Delivery category consists of Original Source operations that generated addresses that were delivered to the DMAF after Census Day. These are:

DSF 4

DSF 5
NRFU
CIFU
New Construction
New Construction and DSF 4
New Construction and DSF 5
DSF 5, New Construction and NRFU
LUCA 1999 Appeals and NRFU
Be Counted Program
Telephone Questionnaire Assistance Program
Be Counted and Telephone Questionnaire Assistance Programs

The timing of a few operations overlapped with questionnaire delivery. When both questionnaire delivery and one of these operations independently added a HU record, the HU record was designated as having been added at the time of Questionnaire Delivery. When one of these overlapping operations added a HU record that was not added during questionnaire delivery, that operation was used to determine whether the HU record was added Pre-Questionnaire Delivery or Post-Questionnaire Delivery. Thus for example, "LUCA 1999 Appeals" is Pre-Questionnaire Delivery, while "LUCA 1999 Appeals and Questionnaire Delivery" is Questionnaire Delivery, and "LUCA 1999 Appeals and NRFU" is Post-Questionnaire Delivery, since NRFU occurred after the time of questionnaire delivery.

Certain HU records have inconsistent data and were not able to be coded for an Original Source. These records have undetermined Original Source and are described as Operation Undetermined in the discussion that follows.

In the table below we give preliminary numbers for the classification of HU records on the DMAF by time of delivery to the DMAF, nationally and by region. This is a classification of all HU records that were ever delivered to the DMAF, as represented on the November 2000 MAF extracts used to determine Tabulation Geography. Some of the HU records on this file were deleted by Census processes, thus the DMAF-deliverable Census 2000 HU count is larger than the final Census 2000 HU count and is not directly comparable to the final HU count from the 1990 Census.

Another limitation to these counts is that certain HU records on the MAF and DMAF were found to be duplicates of each other, although they were not originally identified as such. These records were merged in such a way that the records still exist on the file, but an ID field flag on the record indicates that it is a duplicate of the other identified record. The flag identifying if the record had ever been delivered from the MAF to the DMAF was set back to "N" for the merged records to indicate that the record was no longer deliverable to the DMAF. For the purposes of the tallies used in these evaluations, the actual count of records ever delivered to the DMAF contains some percent of these merged records. The counts in this section should include all HU records that were ever delivered to the DMAF (deliverability flag set to Y). In order to include the merged records, it is also necessary to count the records with valid duplicate IDs, even though not all of these merged records were delivered to the DMAF. The MAF extracts were used for the categorization of these HU records because they contain the values of the Original Source variable, which was used to classify the records as Pre-Questionnaire Delivery, Questionnaire Delivery or Post-Questionnaire Delivery. The Original Source exists only on the data sets created from the MAF extracts.

One further limitation to these counts is that on some HU records the state code changes from one delivery to another, due to updating operations. These numbers were calculated from state-level files that depended on the original state code in one field, but the state code could be different because of these changes. Further research is being done to count and classify these records.

The tables in the appendices classify the addresses delivered to the DMAF according to relationship to the questionnaire delivery and to TEA for each state. As with the Original Source determination, there are also some HU records with undetermined TEA. States with higher percentages of addresses involved in hand delivery of questionnaires should have correspondingly higher percentages of addresses added at the time of Questionnaire Delivery. In particular, we expect that those states with high proportions of List/Enumerate (or Remote Alaska) have a higher percentage of addresses added later than Pre-Questionnaire Delivery.

Census 2000 address list housing unit records, by time of delivery
National and Regional Data
Preliminary Data

Geography	Total	Percent of IDs Added to the Address List from			
		Pre-Questionnaire Delivery Operations	Questionnaire Delivery	Post-Questionnaire Delivery Operations	Operation Undetermined
National	128,691,771	96.7	1.8	1.3	0.3

Region	Northeast	24,545,009	96.7	1.7	1.4	0.3
	South	47,344,579	96.3	2.1	1.4	0.3
	Midwest	29,750,345	97.6	1.0	1.2	0.3
	West	27,051,838	96.4	1.9	1.3	0.4

- Because the address list-building processes for Census 2000 were new, there is no data from the 1990 Census to use for comparisons.

- The national percent of addresses in the Pre-Questionnaire Delivery is 96.7 percent. Regional values do not vary much from the national average. The state values (see Attachment) range from 80.8 to 98.7 percent. The states with the lowest values, ranging from 80.8 to 92.8 percent, respectively, are: Wyoming, Vermont, Alaska, Maine, New Hampshire, Montana, New Mexico, and Nevada. The states with the highest values of Pre-Questionnaire Delivery, from 97.5 to 98.7 percent, respectively, are: Kansas, Indiana, Illinois, Pennsylvania, Michigan, Delaware, California, Massachusetts, Ohio, Connecticut, New Jersey, Rhode Island, Maryland, District of Columbia.

- The national level of addresses added during Questionnaire Delivery is 1.8 percent. This reflects areas of List/Enumerate operations, as well as the operations in which HU records were added at the time of questionnaire delivery, which were Update/Leave and Update/Enumerate.

- There is a regional disparity in percentage of HU records added during Questionnaire Delivery, with the Midwest region significantly lower and the South region significantly higher. One possible reason for this is that the southern states had more territory in the questionnaire delivery operations, while the midwest had less, which is not known at this time. The more likely reason for this is that population and housing are growing quickly in southern states but less quickly than the national average in the Midwest.

- The values of Post-Questionnaire Delivery and Undetermined Original Source for all states are negligible, as well as nearly equal. Thus the same states with lower percentages of Pre-Questionnaire Delivery have higher percentages of Questionnaire Delivery. The state values of Questionnaire Delivery range from 0.1 to 17.8 percent.

- The attachment on TEA distributions gives additional explanatory data for high rates of addresses added in Questionnaire Delivery operations. Update/Leave areas contained 18.8 percent of the total HU records, with state values ranging from 0.1 to 72.4 percent. The Update/Enumerate operation accounted for 0.8 percent of the HU records nationally, but state values ranged from 0.0 to 8.4 percent. The List/Enumerate areas, for which almost all records were added during questionnaire delivery, have 0.3 percent of the HU records nationally, with state values ranging from 0.0 to 15.8 percent.

B. Description of the Quality Assurance Program

The QA program for Address List Development had the following three objectives:

- Prevent errors due to lack of knowledge or understanding on the part of the lister/enumerator.
- Control coverage and content errors.
- Promote continuous improvement of performance.

These objectives were met by using a combination of the following four tools. Not every tool was used to meet each objective.

- Initial Observation. A crew leader (CL) or crew leader assistant (CLA) conducted an initial observation to ensure the listers/enumerators produced work according to the established procedures. The crew leader or assistant observed the listers/enumerators working in the field. If they found any errors, the crew leader or assistant informed the lister/enumerator of the errors and retrained the lister/enumerator.

- Dependent Review. Following the completion of (or throughout) an Assignment Area (AA), a CL or CLA checked a random sample of the completed work. The CL recorded each housing unit (HU) sampled and the type of error(s) identified, if any, to determine whether it was acceptable. If the number of errors in the sample was above the tolerance level, the AA was rejected and recanvassed. If the number of errors in the sample was below the tolerance level, the AA was accepted. Accepted AAs continued to the field office. The CL or CLA also informed the lister/enumerator of errors made and retrained the lister/enumerator as necessary.

- Reinterview (List/Enumerate (L/E) and Update/Enumerate (U/E) only). A separate office staff conducted a review of enumerators' work to ensure accuracy of data collection. Throughout the operation, the Operations Control System (OCS2000) selected cases administratively based upon a comparison of each enumerator's data to the data in the crew leader district. If an enumerator's data were out of tolerance for the crew leader district, the supervisor entered the enumerator's name into the OCS200 and the system began selecting cases for reinterview. Clerks transcribed original information onto a reinterview form for selected reinterview cases. Reinterview staff contacted households by telephone or personal visit to conduct the reinterview. A supervisor reviewed the reinterview results, decided if falsification existed, and took the appropriate action. A Field Operations Supervisor (FOS) notified enumerators of performance errors or discrepant results.

- <u>Office Review.</u> An office clerk performed the office review by reviewing 100 percent of housing units listed in the address binders. When an address binder did not meet the acceptable quality level, it was returned to the enumerator for corrections.

The following preliminary results are available from the Address List Development program:

- QA coverage (work assignments reviewed) ranged from 56.8 percent to 80.5 percent. The expected QA coverage range was between 75 percent and 100 percent.

- Less than four percent of the work assignments failed the QA checks. We expected no more than five percent to fail.

- The reinterview workload was one percent for the U/E operation and two percent for the L/E operation. We expected approximately one percent of the cases completed by the operation would be selected for reinterview.

- Discrepant result were found in approximately 10 percent of the U/E reinterview cases and approximately six percent of the L/E reinterview cases.[1]

[1] These data have undergone an initial round of edits, but further edits are expected. These results should be considered very preliminary.

II. Processing

A. Preliminary Results of the Census 2000 Housing Unit ID Inventory Processing

The data for this section come from two sources, the Decennial Master Address File (DMAF) and the Hundred percent Census Edited File with the reinstated housing unit IDs included (HCEF_D'). For the *Total* column the numbers were generated using the DMAF. For the *In Census 2000* column the numbers were generated using the HCEF_D'. The number of housing unit IDs removed from Census 2000 was determined by subtracting the number of housing unit IDs in Census 2000 from the total of housing unit IDs.

Housing units were removed from the census process from one of three activities. During the first activity housing units were removed if two independent census operations determined the housing unit not to exist and there was no data capture or if two addresses were determined (matched) to be the same housing unit. The census operations which were involved are block canvassing, questionnaire mailing, questionnaire delivery, nonresponse followup, coverage improvement followup and field verification. The second activity removed some housing units when there was conflicting information concerning the existence of the housing unit; either nonresponse followup or coverage improvement followup determined the housing unit did not exist, however a data capture existed for the housing unit. Rules were established to determine the final status of the housing unit in these cases. Finally, the third activity identified address duplication through a set of expanded address and person matching rules. The housing units identified as duplicates from this process were removed from the census.

The table below contains information on the number and percent of housing units determined not to exist and thus were removed for the nation and the four regions. See the attachment for state level data.

Number and Percent of Housing Unit IDs Determined to Not Exist
Housing Unit IDs that removed from Census 2000
National and Regional Data

Geography		Housing Unit IDs			
		Total	In Census 2000	Removed from Census 2000	
				Number	Percent
National		126,276,807	115,904,641	10,372,166	8.2%
Region	Northeast	24,260,015	22,180,440	2,079,575	8.6%
	South	46,216,140	42,382,546	3,833,594	8.3%
	Midwest	29,305,631	26,963,635	2,341,996	8.0%
	West	26,495,021	24,378,020	2,117,001	8.0%

- Data from 1990 are not available
- Nationally 8.2 percent of the housing units in the DMAF were determined not to exist and thus removed
- Regionally between 8.0 and 8.6 percent of the housing units in the DMAF were determined not to exist and thus removed
- At the state level, the percent of housing units in the DMAF that were determined not to exist and thus removed ranged from 5.4 to 16.1 percent
- States with the smallest percent of addresses determined not to exist and removed were Nebraska (5.4 percent), Virginia (5.5 percent), Nevada (5.6 percent), Iowa (5.8 percent) and South Dakota (6.1 percent)
- States with the largest percent of addresses determined not to exist and removed were Lousiana (10.7 percent), Georgia (11.5 percent), Illinois (12.0 percent), South Carolina (12.5 percent) and Hawaii (16.1 percent)

B. Primary Selection Algorithm

The Primary Selection Algorithm (PSA) is applied to a defined subset of response records that have been assigned housing unit (HU) IDs. The purpose of the PSA is to select return and person records that may be included on census files defined by subsequent processes.

More than one response to the census may be received for a given address. This occurs because there are several ways to respond to the census. A person may mail back the census form delivered to his home; he may be interviewed by a census enumerator; he may fill in a Be Counted Form and mail it in; he may fill out a form online and return it via the Internet; he may be enumerated at a group quarters (e.g., a military base) and elect to fill in a different address (i.e., Usual Home Elsewhere (UHE)) at which he thinks he should be counted. Each of these types of responses that arrive for the same housing unit address will create a return coded to the same census ID. It is the job of the PSA to analyze these responses and select from among them the records that it deems most likely to represent the actual census household.

There are two main categories of returns. Standard returns includes mail returns, enumerator returns, internet returns, and CATI returns. These returns all have census provided information on them which identifies the address the return should enumerate. Other returns such as Be Counted Forms or enumerator returns not pre-printed with address information used for the enumeration of persons who were living at a different address on Census Day or who usually live at a different address other than the one the enumerator visited are called Respondent Provided Address (RPA) returns. There are two types of RPAs; whole household RPAs list all persons in the household while partial household RPAs list one person or more than one person but not the entire household.

PSA processing is performed one census housing unit at a time. Within each census housing unit, returns with one or more persons in common are combined to form a single PSA Household. Returns that are identified as vacant are combined into one PSA household. If more than one PSA household exists, one household is selected to represent the census housing unit based on a set of criteria. In some instances, person records from another household consisting of partial RPAs may be added to the selected household.

The objective of the PSA is to select the person and return records that best describe the household that lived at the address on Census Day, i.e, the “census household.” The PSA should, as much as possible, avoid erroneously enumerating or omitting people when more than one form is returned for a Census ID. The benefit of implementing the PSA is a more accurate census count.

1. Number of Returns and Number of PSA Households Per Census Housing Unit

Multiple returns can be received from one census housing unit. This table shows that a housing unit returned two or more returns 9.46 percent of the time.

Census Returns Per Census Housing Unit

Number of returns	**Number of housing units (Percent of total)**
1	107,305,027 (90.54)
2	10,740,311 (9.06)
3+	473,635 (0.40)
Total	118,518,973

2. Census Returns Per PSA Household

A PSA household may consist of more than one return. When more than one return is present in a household, PSA designates one return as the "basic" return according to a set of rules. The remaining returns in the PSA household are referred to as "other" returns. Not all census returns are eligible for PSA. Blank returns, enumerator replacement forms and returns for deleted housing units are ineligible to be placed into a PSA household. There were 130,267,656 total census returns of which 2,656,951 were ineligible for PSA.

The table below shows the number of census housing units with one, two, and three or more PSA household(s) and the number of returns comprising these households. Only one PSA household was formed more than 78 percent of the time a census housing unit had more than one census returns.

When there were no eligible returns for a housing unit, no PSA household was formed. This occurred in 0.13 percent of the census housing units.

Within housing unit with two or more returns there was a total of 22,962,629 census returns of which 13,657,945 were designated as a "basic" return, 6,782,316 were designated as an "other" return and 2,522,368 were ineligible for PSA.

Census Returns Per PSA Household

Number of PSA households	**Total housing units (Percent of total)**	**Total housing units with... (Percent of column total)**		
		One return	**Two returns**	**Three or more returns**
0	158,530 (0.13)	134,583 (0.13)	22,976 (0.21)	971 (0.21)
1	115,964,314 (97.85)	107,170,444 (99.87)	8,549,216 (79.60)	244,654 (51.65)
2	2,349,988 (1.98)		2,168,119 (20.19)	181,869 (38.40)
3+	46,141 (0.04)			46,141 (9.74)
Total	118,518,973	107,305,027	10,740,311	473,635

3. Duplicate Returns

When there are at least two returns in a household, one “other” return may duplicate persons on the “basic” return. When there is more than one vacant return at an ID, all vacant returns form one PSA household. If all of the persons on an “other” return are on the “basic” return the “other” return it is said to be a duplicate of the “basic” return. If an “other” return has at least one person not listed on the “basic” return, it is not a duplicate return. Vacant and occupied duplicates account for 94.37 percent of all “other” eligible returns.

The table below shows the number of eligible “other” returns by the occupancy status of the PSA household.

Duplicate Returns in PSA Households Comprised of Two or More Returns

Type of “other” return and occupancy status	**Number of “other” returns**	**Percent of all “other” returns**
Vacant Duplicate	2,711,735	39.98
Occupied and Undetermined Status Duplicate*	3,689,141	54.39
Occupied and Undetermined Status Non-Duplicate	381,440	5.63

* The occupancy status could not be determined for small fraction of the PSA households.

4. POP Count Changes as a Result of PSA

The household size of the "basic" return determines the minimum size of the PSA household. Persons from "other" returns in the household may be added under certain conditions. These additions may or may not increase the size of the PSA household.

PSA Effect on Population Counts

Status of PSA household		**Number of PSA households with two or more returns (Percent)**
Occupied Household -	'Other' returns added to household size	295,561 (2.63)
Occupied Household -	No additions from 'other' returns	7,115,082 (63.45)
Vacant Household		3,756,622 (33.50)
Other Type of Household With or Without Addition*		46,681 (0.42)

*Other types of households are those where the occupied or vacant status could not be determined

The average household size of all IDs was 2.43 persons. The average increase per household was 0.04 persons.

5. Types of PSA Households

The tables below categorize PSA households into four main types: 1) **occupied** PSA households that are not RPAs, 2) **vacant** PSA households, 3) **whole household RPAs,** and 4) **partial household RPAs**. RPAs include returns such as Be Counted Forms or enumerator returns not pre-printed with address information used for the enumeration of persons who were living at a different address on Census Day or who usually live at a different address other than the one the enumerator visited. The category type into which each PSA household is placed is determined by the "basic" form for the PSA household.

At housing units where we have two PSA households, this table shows the number of census housing units with each of several combinations of these PSA household types for those housing unit with two PSA households.

Number of Census Housing Units with Two PSA Households by Combination of PSA Household Types

Combination of PSA household types	**Number (Percent of census housing units)**
Occupied/Occupied	899,060 (38.26)
Occupied/Vacant	1,056,385 (44.95)
Occupied/Whole Household RPA	94,143 (4.01)
Vacant/Whole Household RPA	35,240 (1.50)
Occupied/Partial Household RPA	79,255 (3.37)
Vacant/Partial Household RPA	10,216 (0.43)
All Other Combinations	175,689 (7.48)

This table shows how often the vacant household was selected by PSA over the occupied or Whole Household RPA household within the categories of Occupied/Vacant and Vacant/Whole Household RPA.

Housing Status Chosen When a Census Housing Unit Consists of Two PSA Households; one Occupied and one Vacant

Combination of PSA household types	**Number of times the vacant household was selected by PSA (Percent of housing units)**
Occupied/Vacant	62,255 (5.89)
Vacant/Whole Household RPA	9,438 (26.78)

III. Completeness of the Data - Item Imputation - Housing Units Only

The following preliminary imputation rates consider all cases that were edited, allocated, or substituted according to the Hundred Percent Census Edited File with the reinstated housing unit IDs included (HCEF_D'). The universe for this analysis was restricted to housing units included in the Census and persons associated with those housing units. Each housing unit and person record contained a form type variable which was used to determine whether the form of the record was self-administered or enumerator-administered. Self-administered forms are filled out by someone within the housing unit. The different types of self-administered forms include the short and long forms used for mailout/mailback, the short and long forms for update/leave, and BeCounted forms. Enumerator-administered forms are forms filled out by a Census enumerator. The form types include the short and long forms for enumerators and the enumerator supplements. Form types that were not logical for this analysis were ignored. These included forms that were used for group quarters enumeration purposes: Individual Census Questionnaires (short and long), Individual Census Reports (short and long), Military Census Reports, and Shipboard Census Reports. These forms included a small number and percent of persons.

The imputation rates for the five items below use the allocation flag variables on the housing unit and person records. Three different types of imputation can occur on each record: edit, allocation, or substitution. An edit is performed when a response for a data item is either missing or not consistent to other responses, and an item value can be determined based on provided information from that same person. Allocations, or computer assignments of acceptable codes in place of unacceptable entries or blanks, are needed most often when an entry for a given items is lacking or when the information reported for a person or housing unit on that item is inconsistent with other information for that same person or housing unit. This is done by grabbing a response from another person within the household or from a person in a nearby household. A substitution occurs when a full set of characteristics for a person or housing unit needs to be assigned. This happens because a questionnaire contains no information for the household and/or no information for the people within the household. A nearby housing unit with complete information is selected as a substitute and the responses are used to fill the missing data items. This housing unit is selected using a nearest neighbor hot deck.

If the response to an item was unchanged through these imputation procedures, it remained a reported value. However, if the response was modified by editing, allocating, or substituting, then the response was considered to have an imputed value. An imputation rate is then computed by tallying the number of imputed cases and dividing it by the total number of reported and imputed cases combined.

The "Total" column in each table represents the overall imputation rate for each specific item. "Self-Administered" in the tables below refers to imputation rates for only self-administered forms. Similarly, “Enumerator-Administered” in the tables refers to imputation rates for only enumerator-administered forms. The “Difference” column refers to the self-administered imputation rate subtracted from the enumerator-administered imputation rate.

Due to the fact that no comparable numbers exist, 1990 imputation rates for the five items below are not provided.

A. Preliminary Results for Age

Imputation Rates for Age
National and Regional Data
Preliminary Data

Geography		Total	Self-Administered	Enumerator-Administered	Difference
National		7.2	4.5	15.4	10.9
Region	Northeast	7.6	4.5	16.7	12.2
	South	7.5	4.5	15.6	11.1
	Midwest	5.9	3.6	14.7	11.1
	West	7.7	5.4	14.5	9.1

- The national imputation rate (total) for the age characteristic is 7.2 percent. The self-administered imputation rate is 4.5 percent and the enumerator-administered imputation rate is 15.4 percent. This creates a difference of 10.9 percentage points between enumerator-administered and self-administered rates.

- The Midwest Region has the lowest total (5.9 percent) and self-administered (14.7 percent) imputation rates for age. The Northeast Region carries the highest enumerator-administered imputation rate (16.7 percent), and this causes it to have the highest difference (12.2 percentage points) among the four regions. Similarly, the West Region has the largest self-administered imputation rate (5.4 percent) which causes the smallest difference (9.1 percentage points) among the regions.

- Range for Total (states): 4.6 to 12.2 percent

<u>Lowest:</u>	<u>Highest:</u>
North Dakota (4.6%)	District of Columbia (12.2%)
Iowa (4.8%)	New York (9.2%)
Nebraska (4.8%)	Nevada (8.8%)

- Range for Self-Administered (states): 2.9 to 7.0 percent

  Lowest:
  North Dakota (2.9%)
  Wyoming (3.1%)
  South Dakota (3.1%)
  Wisconsin (3.1%)
  Iowa (3.1%)

  Highest:
  District of Columbia (7.0%)
  California (6.4%)
  Hawaii (5.8%)

- Range for Enumerator-Administered (states): 10.2 to 23.9 percent

  Lowest:
  West Virginia (10.2%)
  North Dakota (10.2%)
  South Dakota (10.9%)
  Alaska (10.9%)

  Highest:
  District of Columbia (23.9%)
  Delaware (21.4%)
  Maryland (19.6%)

- Range for Difference (states): 6.2 to 17.4 percentage points

  Lowest:
  West Virginia (6.2%)
  Alaska (7.2%)
  Utah (7.3%)
  North Dakota (7.3%)

  Highest:
  Delaware (17.4%)
  District of Columbia (16.9%)
  Maryland (15.3%)

- Findings: In all geographies (national, regional, state), the self-administered imputation rates are much lower than the enumerator-administered imputation rates for age.

B. Preliminary Results for Sex

Imputation Rates for Sex
National and Regional Data
Preliminary Data

Geography		Total	Self-Administered	Enumerator-Administered	Difference
National		3.0	1.8	6.6	4.8
Region	Northeast	3.2	1.7	7.7	6.0
	South	3.1	1.8	6.6	4.8
	Midwest	2.4	1.4	6.1	4.7
	West	3.4	2.5	6.3	3.8

- The national imputation rate (total) for the sex characteristic is 3.0 percent. The self-administered imputation rate is 1.8 percent and the enumerator-administered imputation rate is 6.6 percent. This creates a difference of 4.8 percentage points between enumerator-administered and self-administered rates.

- The Midwest Region has the lowest total (2.4 percent), self-administered (1.4 percent), and enumerator-administered (6.1 percent) imputation rates of the four regions for the sex characteristic. The West Region has the highest total (3.4 percent) and self-administered (2.5 percent) rates. This high self-administered imputation rate helps create the smallest rate difference (3.8 percentage points) for the West compared to the other three regions. The Northeast Region has the highest enumerator-administered (7.7 percent) imputation rate. The Northeast also has the largest difference (6.0 percentage points) because of its relatively average self-administered rate.

- Range for Total (states): 1.7 to 5.3 percent

Lowest:	Highest:
North Dakota (1.7%)	District of Columbia (5.3%)
Iowa (1.7%)	New York (4.3%)
Nebraska (1.8%)	Arizona (3.9%)
West Virginia (1.8%)	Nevada (3.9%)

- Range for Self-Administered (states): 1.1 to 3.0 percent

Lowest:	Highest:
North Dakota (1.1%)	California (3.0%)
Iowa (1.1%)	District of Columbia (2.8%)
	Hawaii (2.6%)

- Range for Enumerator-Administered (states)s: 2.7 to 11.2 percent

Lowest:	Highest:
West Virginia (2.7%)	District of Columbia (11.2%)
Maine (3.4%)	Delaware (10.9%)
North Dakota (3.6%)	New York (9.8%)
	Maryland (9.8%)

- Range for Difference (states): 1.2 to 9.5 percentage points

Lowest:	Highest:
West Virginia (1.2%)	Delaware (9.5%)
Maine (2.1%)	District of Columbia (8.4%)
Mississippi (2.1%)	Maryland (8.2%)

- Findings: In all geographies (national, regional, state), the self-administered imputation rates are much lower than the enumerator-administered imputation rates for sex.

C. Preliminary Results for Race

Imputation Rates for Race
National and Regional Data
Preliminary Data

Geography		Total	Self-Administered	Enumerator-Administered	Difference
National		5.0	4.1	7.5	3.4
Region	Northeast	5.0	3.7	8.7	5.0
	South	4.3	3.3	7.1	3.8
	Midwest	3.3	2.4	6.7	4.3
	West	7.7	7.7	7.7	0.0

- The national imputation rate (total) for the race characteristic is 5.0 percent. The self-administered imputation rate is 4.1 percent and the enumerator-administered imputation rate is 7.5 percent. This creates a difference of 3.4 percentage points between enumerator-administered and self-administered rates.

- The Midwest Region has the lowest total (3.3 percent), self-administered (2.4 percent), and enumerator-administered (6.7 percent) imputation rates for race. By far, the West has the highest total (7.7 percent) imputation rate of the four regions. The West Region also has the highest self-administered (7.7 percent) rate, which is the same as its enumerator-administered rate, thus creating a difference of 0.0 percentage points. The Northeast Region has the highest enumerator-administered (8.7 percent) imputation rate. This causes the Northeast to have the largest rate difference (5.0 percentage points).

- Range for Total (states): 1.9 to 10.0 percent

Lowest:	Highest:
West Virginia (1.9%)	New Mexico (10.0%)
North Dakota (2.2%)	California (9.2%)
Kentucky (2.2%)	Arizona (8.2%)

- Range for Self-Administered (states): 1.5 to 10.5 percent

Lowest:	Highest:

North Dakota (1.5%) | New Mexico (10.5%)
Kentucky (1.6%) | California (9.8%)
West Virginia (1.6%) | Arizona (7.3%)
South Dakota (1.6%)

- Range for Enumerator-Administered (states): 2.6 to 11.9 percent

<u>Lowest:</u>
West Virginia (2.6%)
Kentucky (3.9%)
Mississippi (4.1%)

<u>Highest:</u>
District of Columbia (11.9%)
Delaware (11.4%)
New York (10.8%)

- Range for Difference (states): -2.5 to 9.0 percentage points

<u>Lowest:</u>
California (-2.5%)
New Mexico (-1.5%)
West Virginia (1.0%)

<u>Highest:</u>
Delaware (9.0%)
District of Columbia (8.0%)
Maryland (7.8%)

- Findings: In the national and regional geographies, self-administered imputation rates are lower than enumerator-administered rates for the race characteristic except in one region (West) where the rates are the same. On a state level, all but two states (California and New Mexico) have self-administered imputation rates that are lower than the enumerator-administered rate.

D. Preliminary Results for Hispanic Origin

Imputation Rates for Hispanic Origin
National and Regional Data
Preliminary Data

Geography		Total	Self-Administered	Enumerator-Administered	Difference
National		5.4	4.6	7.7	3.1
Region	Northeast	5.6	4.5	8.9	4.4
	South	5.7	5.0	7.4	2.4
	Midwest	4.4	3.8	6.9	3.1
	West	5.8	5.1	7.8	2.7

- The national imputation rate (total) for the Hispanic origin characteristic is 5.4 percent. The self-administered imputation rate is 4.6 percent and the enumerator-administered imputation rate is 7.7 percent. This creates a difference of 3.1 percentage points between enumerator-administered and self-administered rates.

- The Midwest Region has the lowest total (4.4 percent), self-administered (3.8 percent), and enumerator-administered (6.9 percent) imputation rates of all four regions for Hispanic origin. The West has the highest total (5.8 percent) and self-administered (5.1 percent) imputation rates. The Northeast Region carries the highest enumerator-administered (8.9 percent) rate, and this translates into the largest difference (4.4 percentage points) of the four regions.

- Range for Total (states): 3.2 to 9.7 percent

Lowest:	Highest:
Iowa (3.2%)	District of Columbia (9.7%)
Nebraska (3.4%)	Hawaii (7.3%)
North Dakota (3.5%)	New York (7.0%)

- Range for Self-Administered (states): 2.9 to 8.3 percent

Lowest:	Highest:
Iowa (2.9%)	District of Columbia (8.3%)
Nebraska (3.0%)	Mississippi (7.5%)
Vermont (3.0%)	Hawaii (7.0%)

- Range for Enumerator-Administered (states): 2.8 to 13.0 percent

Lowest:	Highest:
West Virginia (2.8%)	District of Columbia (13.0%)
Kentucky (4.0%)	Delaware (11.9%)
Maine (4.3%)	New York (11.0%)

- Range for Difference (states): -2.9 to 6.1 percentage points

Lowest:	Highest:
Mississippi (-2.9%)	Delaware (7.8%)
West Virginia (-2.1%)	Maryland (6.3%)
Kentucky (-0.5%)	Indiana (6.1%)
	Arizona (6.1%)

- Findings: In the national and regional geographies, self-administered imputation rates are lower than enumerator-administered rates in every case for the Hispanic origin characteristic. On a state level, all but four states (Mississippi, West Virginia, Kentucky, and Arkansas) have self-administered imputation rates that are lower than the enumerator-administered rate.

E. Preliminary Results for Tenure

Preliminary Imputation Rates for Tenure
National and Regional Data
Preliminary Data

Geography		Total	Self-Administered	Enumerator-Administered	Difference
National		5.3	3.0	12.4	9.4
Region	Northeast	5.7	3.1	13.1	10.0
	South	5.7	3.1	13.1	10.0
	Midwest	4.7	2.9	12.1	9.2
	West	4.8	2.9	10.5	7.6

- The national imputation rate (total) for the tenure characteristic is 5.3 percent. The self-administered imputation rate is 3.0 percent and the enumerator-administered imputation rate is 12.4 percent. This creates a difference of 9.4 percentage points between enumerator-administered and self-administered rates.

- The Midwest Region has the lowest total (4.7 percent) imputation rate of the four regions for tenure. The Midwest, along with the West Region, have the lowest self-administered imputation rates at 2.9 percent. The West also has the lowest enumerator-administered (10.5 percent) imputation rate as well as the smallest rate difference (7.6 percentage points) of the regions. The Northeast and South Regions carry the same imputation rates for all four categories. Each of these rates is the highest among the regions: total (5.7 percent), self-administered (3.1 percent), enumerator-administered (13.1 percent), and difference (10.0 percentage points).

- Range for Total (states): 3.6 to 8.3 percent

Lowest:	Highest:
Alaska (3.6%)	District of Columbia (8.3%)
Utah (3.7%)	Alabama (7.4%)
Ohio (3.9%)	Delaware (6.6%)
	New York (6.6%)

- Range for Self-Administered (states): 2.1 to 4.0 percent

Lowest:	Highest:
Utah (2.1%)	Mississippi (4.0%)
Colorado (2.4%)	Arkansas (3.9%)

	District of Columbia (3.6%)
	Alabama (3.6%)

- Range for Enumerator-Administered (states): 4.7 to 19.4 percent

Lowest:	Highest:
Alaska (4.7%)	Delaware (19.4%)
Oregon (8.5%)	District of Columbia (19.2%)
Utah (8.5%)	Alabama (17.5%)

- Range for Difference (states): 1.9 to 16.9 percentage points

Lowest:	Highest:
Alaska (1.9%)	Delaware (16.9%)
Oregon (5.8%)	District of Columbia (15.6%)
West Virginia (5.9%)	Alabama (13.9%)

- Findings: In all geographies (national, regional, state), the self-administered imputation rates are much lower than the enumerator-administered imputation rates for tenure.

F. Preliminary Findings - Summary

- Of all five characteristics, age has the largest national imputation rate (total) and the largest national difference between enumerator-administered and self-administered rates. Both of these seem to be caused by the extremely high enumerator-administered national imputation rates. This could be due to the fact that the age and date of birth items were included in the same question on the enumerator questionnaire. The enumerator may have only asked for the date of birth information to speed up an interview figuring that the age could be computed from a person's date of birth. In a case where the enumerator forgot or incorrectly filled in the age portion of the question after receiving the date of birth, an edit would occur to correctly fill the age field. If edits would not have been included as a type of imputation for this analysis, the national age imputation rate might be lower.

- For all five characteristics nationally, the self-administered imputation rates are considerably lower than the enumerator-administered rates.

- The Midwest Region has the lowest total imputation rate of the four regions for all five characteristics. This could be attributed to better reporting on the self-administered forms, where the Midwest rates are also ranked as the best in comparison to the other regions.

- In general, a state remains consistent across the five characteristics when compared to the other states. That is, a state does not go from having one of the best (low) imputation rates for one characteristic to having the one of the worst (high) imputation rates for another characteristic.

- It appears that a state with a lower self-administered imputation rate translates into a lower total imputation rate compared to other states.

- When a state has a low enumerator-administered imputation rate, the difference between the self-administered and enumerator-administered rates is also low compared to other states.

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Attachments

Census address list housing unit records, by time of delivery
National and State Data
Preliminary Counts

Geography		Total	Number and Percent of IDs Added to the Address List from							
			Pre-Questionnaire Delivery Operations		Questionnaire Delivery		Post-Questionnaire Delivery Operations		Operation Undetermined	
			#	%	#	%	#	%	#	%
National		128,691,771	124,405,492	96.7	2,316,379	1.8	1,737,311	1.3	381,597	0.3
State	Alabama	2,226,880	2,132,891	95.8	56,913	2.6	33,863	1.5	3,213	0.1
	Alaska	290,803	243,154	83.6	40,035	13.8	6,048	2.1	1,566	0.5
	Arizona	2,442,284	2,304,295	94.4	89,917	3.6	32,504	1.3	16,568	0.7
	Arkansas	1,311,772	1,246,236	95.0	49,131	3.7	13,691	1.0	2,714	0.2
	California	13,413,871	13,136,059	97.9	102,570	0.8	135,847	1.0	39,395	0.3
	Colorado	1,986,641	1,909,776	96.1	46,877	2.4	22,677	1.1	7,311	0.4
	Connecticut	1,517,176	1,489,112	98.2	5,256	0.3	16,450	1.1	6,358	0.4
	Delaware	370,219	362,352	97.9	4,222	1.1	3,059	0.8	586	0.2
	District of Columbia	296,878	293,139	98.7	216	0.1	2,465	0.8	1,058	0.4
	Florida	8,187,877	7,962,003	97.2	79,992	1.0	119,402	1.5	26,480	0.3
	Georgia	3,932,790	3,785,927	96.3	75,213	1.9	62,849	1.6	8,801	0.2
	Hawaii	550,586	534,174	97.0	6,334	1.2	8,224	1.5	1,854	0.3

	Idaho	596,053	563,363	94.5	13,643	2.3	16,362	2.7	2,685	0.5
	Illinois	5,658,489	5,527,505	97.7	24,152	0.4	92,251	1.6	14,581	0.3
	Indiana	2,837,223	2,771,365	97.7	15,195	0.6	43,570	1.5	6,373	0.2
	Iowa	1,328,772	1,292,255	97.3	19,950	1.5	11,969	0.9	4,598	0.3
	Kansas	1,231,192	1,200,987	97.5	15,537	1.3	10,128	0.8	4,540	0.4
	Kentucky	1,945,361	1,865,041	95.9	50,695	2.6	24,428	1.3	5,197	0.3
	Louisiana	2,099,677	2,024,809	96.4	43,540	2.1	27,555	1.3	3,773	0.3
	Maine	709,305	605,174	85.3	96,414	13.6	6,492	0.9	1,225	0.2
	Maryland	2,320,497	2,279,455	98.2	11,621	0.5	24,438	1.1	4,983	0.2
	Massachusetts	2,848,405	2,795,723	98.2	11,117	0.4	29,957	1.1	11,608	0.4
	Michigan	4,614,720	4,518,406	97.9	38,869	0.8	46,038	1.0	11,407	0.2
	Minnesota	2,250,915	2,190,012	97.3	32,011	1.4	21,470	1.0	7,422	0.2
	Mississippi	1,308,752	1,242,006	94.9	36,217	2.8	26,348	2.0	4,181	0.3
	Missouri	2,694,326	2,613,950	97.0	52,139	1.9	20,335	0.8	7,902	0.4
	Montana	452,085	413,885	91.6	31,839	7.0	4,064	0.9	2,297	0.5
	Nebraska	774,108	751,588	97.1	13,276	1.7	5,828	0.8	3,416	0.4
	Nevada	883,053	819,592	92.8	39,772	4.5	19,095	2.2	4,594	0.5
	New Hampshire	591,273	525,388	88.9	57,592	9.7	6,982	1.2	1,311	0.2
	New Jersey	3,605,986	3,540,202	98.2	7,979	0.2	47,795	1.3	10,010	0.3
	New Mexico	880,622	816,500	92.7	49,074	5.6	11,219	1.3	3,829	0.4

	New York	8,651,115	8,342,982	96.4	123,570	1.4	156,730	1.8	27,833	0.3
	North Carolina	3,929,467	3,733,073	95.0	124,465	3.2	62,622	1.6	9,307	0.2
	North Dakota	316,042	300,302	95.0	11,986	3.8	2,314	0.7	1,440	0.5
	Ohio	5,164,457	5,068,491	98.1	24,216	0.5	57,657	1.1	14,093	0.3
	Oklahoma	1,653,495	1,589,931	96.2	43,713	2.6	14,812	0.9	5,039	0.3
	Oregon	1,615,538	1,558,865	96.5	17,996	1.1	29,775	1.8	8,902	0.6
	Pennsylvania	5,800,967	5,672,719	97.8	54,049	0.9	62,182	1.1	12,017	0.2
	Rhode Island	480,124	471,304	98.2	2,612	0.5	5,246	1.1	962	0.2
	South Carolina	2,040,919	1,951,367	95.6	55,188	2.7	29,323	1.4	5,041	0.2
	South Dakota	350,536	330,338	94.2	15,105	4.3	3,475	1.0	1,618	0.5
	Tennessee	2,733,483	2,617,135	95.7	58,628	2.1	48,355	1.8	9,365	0.3
	Texas	8,914,555	8,560,484	96.0	194,945	2.2	133,146	1.5	25,980	0.3
	Utah	960,599	920,122	95.8	24,665	2.6	13,435	1.4	2,377	0.2
	Vermont	340,658	281,224	82.6	54,689	16.1	4,153	1.2	592	0.2
	Virginia	3,156,582	3,058,381	96.9	59,422	1.9	31,098	1.0	7,681	0.2
	Washington	2,734,044	2,662,041	97.4	18,356	0.7	46,177	1.7	7,470	0.3
	West Virginia	915,375	8874,035	95.5	32,281	3.5	8,309	0.9	750	0.1
	Wisconsin	2,529,565	2,457,760	97.2	28,663	1.1	35,462	1.4	7,680	0.3
	Wyoming	245,659	198,614	80.8	43,627	17.8	2,100	0.9	1,318	0.5

Census address list housing unit records, by TEA
National, Regional and State Data
Preliminary Counts

Geography		Total	Number and Percent of IDs by Type of Enumeration Area (TEA)							
			Mailout/Mailback		Update/Leave		Update/Enumerate		List/Enumerate	
			#	%	#	%	#	%	#	%
National		128,691,771	102,975,788	80.0	24,209,899	18.8	1,079,039	0.8	422,533	0.3
Region	Northeast	24,545,009	20,843,333	84.9	3,249,315	13.2	234,816	1.0	217,349	0.9
	South	47,344,579	34,314,575	72.5	12,753,381	26.9	258,087	0.5	15,475	0.0
	Midwest	29,750,345	24,604,228	82.7	4,976,606	16.7	156,800	0.5	11,981	0.0
	West	27,051,838	23,213,652	85.8	3,230,597	11.9	429,336	1.6	177,728	0.7
State	Alabama	2,226,880	1,410,080	63.3	805,427	36.2	11,183	0.5	0	0.0
	Alaska	290,803	172,492	59.3	87,551	30.1	143	0.0	30,606	10.5
	Arizona	2,442,284	1,889,217	77.4	435,681	17.8	91,304	3.7	26,075	1.1
	Arkansas	1,311,772	560,609	43.3	751,096	56.7	0	0.0	0	0.0
	California	13,413,871	12,647,608	94.3	621,540	4.6	108,608	0.8	35,823	0.3
	Colorado	1,986,641	1,448,075	72.9	500,435	25.2	38,097	1.9	0	0.0
	Connecticut	1,517,176	1,419,202	93.5	97,865	6.5	103	0.0	0	0.0
	Delaware	370,219	299,677	80.9	70,540	19.1	0	0.0	0	0.0

	District of Columbia	296,878	296,569	99.9	307	0.1	0	0.0	0	0.0
	Florida	8,187,877	7,439,287	90.9	680,397	8.3	68,016	0.8	0	0.0
	Georgia	3,932,790	2,885,121	73.4	1,047,193	26.6	0	0.0	0	0.0
	Hawaii	550,586	459,813	83.5	90,561	16.4	0	0.0	208	0.0
	Idaho	596,053	471,396	79.1	105,433	17.7	15,762	2.6	3,503	0.6
	Illinois	5,658,489	5,216,579	92.2	441,885	7.8	0	0.0	0	0.0
	Indiana	2,837,223	2,574,052	90.7	255,773	9.0	7,319	0.3	0	0.0
	Iowa	1,328,772	861,298	64.8	467,131	35.2	270	0.0	0	0.0
	Kansas	1,231,192	896,422	72.8	332,012	27.0	2,726	0.2	0	0.0
	Kentucky	1,945,361	1,176,201	60.5	751,877	38.6	16,541	0.9	0	0.0
	Louisiana	2,099,677	1,457,524	69.4	618,691	29.5	23,432	1.1	0	0.0
	Maine	709,305	242,656	34.2	382,758	54.0	8,558	1.2	75,315	10.6
	Maryland	2,320,497	2,107,801	90.8	212,680	9.2	0	0.0	0	0.0
	Massachusetts	2,848,405	2,709,678	95.1	64,499	2.3	74,213	2.6	0	0.0
	Michigan	4,614,720	3,925,949	85.1	688,567	14.9	0	0.0	0	0.0
	Minnesota	2,250,915	1,629,836	72.4	605,334	26.9	15,684	0.7	0	0.0
	Mississippi	1,308,752	804,767	61.5	501,628	38.3	2,134	0.2	0	0.0
	Missouri	2,694,326	1,825,542	67.8	868,706	32.2	0	0.0	0	0.0
	Montana	452,085	120,220	26.6	292,853	64.8	30,203	6.7	8,794	1.9
	Nebraska	774,108	532,258	68.8	233,988	30.2	3,708	0.5	4,025	0.5

	Nevada	883,053	637,566	72.2	201,446	22.8	26,041	2.9	17,989	2.0
	New Hampshire	591,273	331,810	56.1	214,966	36.4	0	0.0	44,490	7.5
	New Jersey	3,605,986	3,491,388	96.8	114,582	3.2	0	0.0	0	0.0
	New Mexico	880,622	475,244	54.0	325,224	36.9	73,669	8.4	6,454	0.7
	New York	8,651,115	7,444,315	86.1	1,098,908	12.7	51,878	0.6	55,949	0.6
	North Carolina	3,929,467	2,136,764	54.4	1,769,097	45.0	23,288	0.6	0	0.0
	North Dakota	316,042	148,157	46.9	146,141	46.2	17,826	5.6	3,913	1.2
	Ohio	5,164,457	4,681,877	90.7	482,566	9.3	0	0.0	0	0.0
	Oklahoma	1,653,495	977,768	59.1	675,669	40.9	0	0.0	0	0.0
	Oregon	1,615,538	1,448,967	89.7	162,227	10.0	2,552	0.2	1,770	0.1
	Pennsylvania	5,800967	4,700,176	81.0	1,002,317	17.3	98,411	1.7	0	0.0
	Rhode Island	480,124	405,297	84.4	73,172	15.2	1,653	0.3	0	0.0
	South Carolina	2,040,919	1,425,636	69.9	607,102	29.7	8,121	0.4	0	0.0
	South Dakota	350,536	163,816	46.7	158,532	45.2	24,141	6.9	4,043	1.2
	Tennessee	2,733,483	1,998,956	73.1	734,442	26.9	0	0.0	0	0.0
	Texas	8,914,555	6,871,834	77.1	1,921,557	21.6	105,342	1.2	15,475	0.2
	Utah	960,599	766,229	79.8	161,535	16.8	25,285	2.6	7,526	0.8
	Vermont	340,658	98,811	29.0	200,248	58.8	0	0.0	41,595	12.2
	Virginia	3,156,582	2,213,654	70.1	942,710	29.9	30	0.0	0	0.0
	Washington	2,734,044	2,547,906	93.2	179,348	6.6	6,642	0.2	86	0.0

	West Virginia	915,375	252,327	27.6	662,968	72.4	0	0.0	0	0.0
	Wisconsin	2,529,565	2,148,442	84.9	295,971	11.7	85,126	3.4	0	0.0
	Wyoming	245,659	128,969	52.5	66,763	27.2	11,030	4.5	38,894	15.8

Mailout/Mailback = TEAs 1 and 6
Update/Leave = TEAs 2, 7 and 9
Update/Enumerate = TEAs 5 and 8
List/Enumerate = TEAs 3 and 4

Note: The complete counts by TEA contain a certain number of address records that have been merged with other address records and now contain no TEA information. These numbers are not shown here but would make the remaining numbers in the table sum to the totals shown.

Number and Percent of Housing Unit IDs Determined Not to Exist
Housing Unit IDs removed from Census 2000
National and State Data
Preliminary Data

		Housing Unit IDs			
				Removed from Census 2000	
Geography		Total	In Census 2000	Number	Percent
National		126,276,807	115,904,641	10,372,166	8.2%
State	Alabama	2,179,657	1,963,711	215,946	9.9%
	Alaska	288,265	260,978	27,287	9.5%
	Arizona	2,417,314	2,189,189	228,125	9.4%
	Arkansas	1,274,650	1,173,043	101,607	8.0%
	California	13,107,542	12,214,549	892,993	6.8%
	Colorado	1,965,768	1,808,037	157,731	8.0%
	Connecticut	1,504,421	1,385,975	118,446	7.9%
	Delaware	368,409	343,072	25,337	6.9%
	District of Columbia	295,182	274,845	20,337	6.9%
	Florida	7,937,571	7,302,947	634,624	8.0%
	Georgia	3,708,750	3,281,737	427,013	11.5%
	Hawaii	548,960	460,542	88,418	16.1%
	Idaho	585,802	527,824	57,978	9.9%
	Illinois	5,552,854	4,885,615	667,239	12.0%
	Indiana	2,794,737	2,532,319	262,418	9.4%
	Iowa	1,309,034	1,232,511	76,523	5.8%
	Kansas	1,210,025	1,131,200	78,825	6.5%
	Kentucky	1,905,170	1,750,926	154,244	8.1%
	Lousiana	2,068,967	1,847,181	221,786	10.7%
	Maine	695,097	651,901	43,196	6.2%
	Maryland	2,292,693	2,145,283	147,410	6.4%
	Massachusetts	2,832,183	2,621,989	210,194	7.4%
	Michigan	4,547,229	4,234,279	312,950	6.9%
	Minnesota	2,211,912	2,065,946	145,966	6.6%
	Mississippi	1,284,940	1,161,953	122,987	9.6%
	Missouri	2,643,651	2,442,017	201,634	7.6%
	Montana	443,108	412,633	30,475	6.9%
	Nebraska	763,849	722,668	41,181	5.4%
	Nevada	876,797	827,457	49,340	5.6%
	New Hampshire	583,474	547,024	36,450	6.2%
	New Jersey	3,579,895	3,310,275	269,620	7.5%
	New Mexico	868,605	780,579	88,026	10.1%
	New York	8,529,607	7,679,307	850,300	10.0%
	North Carolina	3,857,390	3,523,944	333,446	8.6%
	North Dakota	311,631	289,677	21,954	7.0%
	Ohio	5,112,651	4,783,051	329,600	6.4%
	Oklahoma	1,621,526	1,514,400	107,126	6.6%
	Oregon	1,597,106	1,452,709	144,397	9.0%
	Pennsylvania	5,732,579	5,249,750	482,829	8.4%
	Rhode Island	478,179	439,837	38,342	8.0%
	South Carolina	2,003,324	1,753,670	249,654	12.5%
	South Dakota	344,216	323,208	21,008	6.1%
	Tennessee	2,690,789	2,439,444	251,345	9.3%
	Texas	8,751,308	8,157,575	593,733	6.8%
	Utah	854,198	768,594	85,604	10.0%
	Vermont	324,580	294,382	30,198	9.3%
	Virginia	3,071,978	2,904,192	167,786	5.5%
	Washington	2,698,712	2,451,075	247,637	9.2%
	West Virginia	903,836	844,623	59,213	6.6%
	Wisconsin	2,503,842	2,321,144	182,698	7.3%
	Wyoming	242,844	223,854	18,990	7.8%

Imputation Rates for Age
National and State Data
Preliminary Data

Geography	Total	Self-Administered	Enumerator-Administered	Difference
National	7.2	4.5	15.4	10.9
Alabama	8.1	4.5	17.3	12.8
Alaska	6.5	3.7	10.9	7.2
Arizona	8.6	4.8	17.3	12.5
Arkansas	6.6	4.4	12.8	8.4
California	8.3	6.4	14.2	7.8
Colorado	6.7	3.8	16.8	13.0
Connecticut	6.6	4.2	14.6	10.4
Delaware	8.4	4.0	21.4	17.4
District of Columbia	12.2	7.0	23.9	16.9
Florida	7.9	4.6	17.1	12.5
Georgia	8.2	4.9	18.3	13.4
Hawaii	8.6	5.8	15.3	9.5
Idaho	5.8	3.6	12.2	8.6
Illinois	7.6	4.4	18.2	13.8
Indiana	6.6	3.3	17.9	14.6
Iowa	4.8	3.1	12.7	9.6
Kansas	5.7	3.5	13.6	10.1
Kentucky	5.7	3.7	11.5	7.8
Louisiana	7.5	4.6	14.2	9.6
Maine	6.1	3.4	12.1	8.7
Maryland	7.9	4.3	19.6	15.3
Massachusetts	6.7	4.3	14.1	9.8
Michigan	5.6	3.9	13.4	9.5
Minnesota	5.1	3.3	13.9	10.6
Mississippi	7.5	5.3	13.2	7.9
Missouri	5.4	3.6	12.6	9.0
Montana	5.5	3.4	11.4	8.0
Nebraska	4.8	3.2	12.1	8.9
Nevada	8.8	5.0	17.4	12.4
New Hampshire	6.4	3.4	14.9	11.5
New Jersey	7.5	4.6	16.7	12.1
New Mexico	7.9	4.3	14.6	10.3
New York	9.2	5.4	18.8	13.4
North Carolina	7.0	4.3	14.3	10.0
North Dakota	4.6	2.9	10.2	7.3
Ohio	5.3	3.6	12.0	8.4
Oklahoma	6.0	3.9	12.3	8.4
Oregon	6.0	3.9	12.2	8.3
Pennsylvania	6.3	3.8	15.5	11.7
Rhode Island	7.6	4.3	17.3	13.0
South Carolina	7.5	4.4	15.0	10.6
South Dakota	4.9	3.1	10.9	7.8
Tennessee	6.7	4.1	14.0	9.9
Texas	8.1	4.9	15.9	11.0
Utah	5.9	4.1	11.4	7.3
Vermont	6.7	3.3	13.6	10.3
Virginia	6.1	3.9	14.1	10.2
Washington	6.7	3.9	14.4	10.5
West Virginia	5.5	4.0	10.2	6.2
Wisconsin	5.3	3.1	15.7	12.6
Wyoming	6.4	3.1	12.0	8.9

Imputation Rates for Sex
National and State Data
Preliminary Data

Geography	Total	Self-Administered	Enumerator-Administered	Difference
National	3.0	1.8	6.6	4.8
Alabama	3.4	1.7	7.7	6.0
Alaska	2.9	1.4	5.1	3.7
Arizona	3.9	2.0	8.3	6.3
Arkansas	2.4	1.7	4.6	2.9
California	3.8	3.0	6.1	3.1
Colorado	2.9	1.6	7.3	5.7
Connecticut	2.5	1.4	6.0	4.6
Delaware	3.8	1.4	10.9	9.5
District of Columbia	5.3	2.8	11.2	8.4
Florida	3.1	1.8	6.8	5.0
Georgia	3.5	2.0	8.0	6.0
Hawaii	3.7	2.6	6.4	3.8
Idaho	2.4	1.5	5.1	3.6
Illinois	3.4	1.8	8.5	6.7
Indiana	2.9	1.3	8.7	7.4
Iowa	1.7	1.1	4.3	3.2
Kansas	2.0	1.3	4.6	3.3
Kentucky	2.0	1.4	3.8	2.4
Louisiana	3.1	1.8	6.0	4.2
Maine	2.0	1.3	3.4	2.1
Maryland	3.5	1.6	9.8	8.2
Massachusetts	2.5	1.5	5.6	4.1
Michigan	2.2	1.5	5.0	3.5
Minnesota	2.0	1.3	5.3	4.0
Mississippi	2.8	2.2	4.3	2.1
Missouri	2.0	1.4	4.5	3.1
Montana	2.2	1.4	4.3	2.9
Nebraska	1.8	1.2	4.4	3.2
Nevada	3.9	2.2	7.9	5.7
New Hampshire	2.7	1.2	6.7	5.5
New Jersey	3.1	1.7	7.4	5.7
New Mexico	3.6	1.8	7.1	5.3
New York	4.3	2.1	9.8	7.7
North Carolina	2.6	1.7	5.3	3.6
North Dakota	1.7	1.1	3.6	2.5
Ohio	1.9	1.3	4.2	2.9
Oklahoma	2.1	1.5	4.1	2.6
Oregon	2.3	1.5	4.7	3.2
Pennsylvania	2.4	1.4	6.2	4.8
Rhode Island	3.2	1.5	8.2	6.7
South Carolina	3.2	1.7	6.8	5.1
South Dakota	1.9	1.2	4.3	3.1
Tennessee	2.5	1.5	5.3	3.8
Texas	3.6	2.1	7.3	5.2
Utah	2.5	1.8	4.7	2.9
Vermont	2.7	1.2	5.7	4.5
Virginia	2.5	1.5	6.0	4.5
Washington	2.7	1.6	5.6	4.0
West Virginia	1.8	1.5	2.7	1.2
Wisconsin	2.2	1.2	7.0	5.8
Wyoming	3.0	1.2	6.1	4.9

Imputation Rates for Race
National and State Data
Preliminary Data

Geography	Total	Self-Administered	Enumerator-Administered	Difference
National	5.0	4.1	7.5	3.4
Alabama	3.5	1.9	7.8	5.9
Alaska	4.1	3.1	5.8	2.7
Arizona	8.2	7.3	10.3	3.0
Arkansas	2.9	2.3	4.7	2.4
California	9.2	9.8	7.3	-2.5
Colorado	6.4	5.4	9.7	4.3
Connecticut	4.4	3.6	7.3	3.7
Delaware	4.7	2.4	11.4	9.0
District of Columbia	6.3	3.9	11.9	8.0
Florida	4.2	3.0	7.5	4.5
Georgia	4.2	2.8	8.3	5.5
Hawaii	5.9	5.4	7.2	1.8
Idaho	4.2	3.5	6.2	2.7
Illinois	5.3	4.1	9.2	5.1
Indiana	3.7	2.0	9.4	7.4
Iowa	2.4	1.9	5.0	3.1
Kansas	3.5	2.9	5.7	2.8
Kentucky	2.2	1.6	3.9	2.3
Louisiana	3.2	2.1	5.8	3.7
Maine	2.5	1.7	4.2	2.5
Maryland	4.3	2.5	10.3	7.8
Massachusetts	4.2	3.3	7.1	3.8
Michigan	2.9	2.3	5.6	3.3
Minnesota	2.8	2.1	6.5	4.4
Mississippi	2.7	2.2	4.1	1.9
Missouri	2.5	1.9	4.7	2.8
Montana	3.1	2.2	5.4	3.2
Nebraska	2.8	2.3	4.8	2.5
Nevada	7.4	6.5	9.3	2.8
New Hampshire	3.6	2.1	8.1	6.0
New Jersey	5.1	4.1	8.2	4.1
New Mexico	10.0	10.5	9.0	-1.5
New York	6.7	5.1	10.8	5.7
North Carolina	3.2	2.3	5.6	3.3
North Dakota	2.2	1.5	4.4	2.9
Ohio	2.3	1.8	4.3	2.5
Oklahoma	3.2	2.5	4.9	2.4
Oregon	4.1	3.5	6.0	2.5
Pennsylvania	3.3	2.4	6.8	4.4
Rhode Island	4.8	3.3	9.3	6.0
South Carolina	3.4	2.0	6.8	4.8
South Dakota	2.3	1.6	4.5	2.9
Tennessee	2.8	1.8	5.5	3.7
Texas	7.3	6.7	8.6	1.9
Utah	4.3	3.8	5.7	1.9
Vermont	3.2	1.7	6.5	4.8
Virginia	3.3	2.3	6.9	4.6
Washington	4.8	3.9	7.2	3.3
West Virginia	1.9	1.6	2.6	1.0
Wisconsin	3.3	2.2	8.1	5.9
Wyoming	4.5	3.0	7.1	4.1

Imputation Rates for Hispanic Origin
National and State Data
Preliminary Data

Geography	Total	Self-Administered	Enumerator-Administered	Difference
National	5.4	4.6	7.7	3.1
Alabama	6.5	5.8	8.4	2.6
Alaska	4.7	3.9	5.9	2.0
Arizona	6.0	4.1	10.2	6.1
Arkansas	5.2	5.2	4.9	-0.3
California	6.2	5.9	7.4	1.5
Colorado	5.1	3.8	9.7	5.9
Connecticut	4.6	3.8	7.3	3.5
Delaware	6.1	4.1	11.9	7.8
District of Columbia	9.7	8.3	13.0	4.7
Florida	5.3	4.4	8.0	3.6
Georgia	6.6	5.9	8.6	2.7
Hawaii	7.3	7.0	8.1	1.1
Idaho	3.9	3.2	6.2	3.0
Illinois	5.5	4.3	9.3	5.0
Indiana	4.9	3.5	9.6	6.1
Iowa	3.2	2.9	5.0	2.1
Kansas	3.9	3.3	5.9	2.6
Kentucky	4.4	4.5	4.0	-0.5
Louisiana	6.1	6.0	6.2	0.2
Maine	3.6	3.4	4.3	0.9
Maryland	6.0	4.6	10.9	6.3
Massachusetts	4.8	3.9	7.3	3.4
Michigan	4.5	4.2	6.0	1.8
Minnesota	3.8	3.2	6.7	3.5
Mississippi	6.7	7.5	4.6	-2.9
Missouri	4.1	4.0	4.9	0.9
Montana	4.2	3.6	5.6	2.0
Nebraska	3.4	3.0	5.0	2.0
Nevada	6.1	4.6	9.5	4.9
New Hampshire	4.4	3.1	8.2	5.1
New Jersey	5.4	4.4	8.4	4.0
New Mexico	6.5	4.9	9.3	4.4
New York	7.0	5.4	11.0	5.6
North Carolina	5.3	5.1	5.8	0.7
North Dakota	3.5	3.2	4.7	1.5
Ohio	4.0	3.9	4.6	0.7
Oklahoma	4.5	4.2	5.2	1.0
Oregon	4.1	3.5	6.0	2.5
Pennsylvania	4.6	4.0	7.1	3.1
Rhode Island	5.3	3.8	9.4	5.6
South Carolina	6.3	6.0	7.3	1.3
South Dakota	3.6	3.2	5.2	2.0
Tennessee	5.1	4.9	5.7	0.8
Texas	5.9	4.8	8.6	3.8
Utah	4.0	3.4	5.9	2.5
Vermont	4.2	3.0	6.7	3.7
Virginia	5.2	4.7	7.2	2.5
Washington	5.0	4.2	7.3	3.1
West Virginia	4.4	4.9	2.8	-2.1
Wisconsin	4.1	3.2	8.3	5.1
Wyoming	4.6	3.1	7.2	4.1

Imputation Rates for Tenure
National and State Data
Preliminary Data

Geography	Total	Self-Administered	Enumerator-Administered	Difference
National	5.3	3.0	12.4	9.4
Alabama	7.4	3.6	17.5	13.9
Alaska	3.6	2.8	4.7	1.9
Arizona	6.1	2.8	14.5	11.7
Arkansas	6.0	3.9	12.4	8.5
California	4.7	3.1	10.0	6.9
Colorado	4.7	2.4	12.5	10.1
Connecticut	4.7	2.8	11.4	8.6
Delaware	6.6	2.5	19.4	16.9
District of Columbia	8.3	3.6	19.2	15.6
Florida	5.6	2.8	13.9	11.1
Georgia	6.4	3.3	15.5	12.2
Hawaii	4.6	2.6	9.6	7.0
Idaho	4.5	2.6	9.8	7.2
Illinois	5.5	2.9	14.3	11.4
Indiana	5.6	2.6	16.3	13.7
Iowa	4.3	2.9	11.2	8.3
Kansas	4.7	2.8	11.9	9.1
Kentucky	4.7	3.0	9.8	6.8
Louisiana	6.1	3.3	12.7	9.4
Maine	5.6	2.9	11.5	8.6
Maryland	5.7	2.5	16.2	13.7
Massachusetts	4.9	3.0	10.5	7.5
Michigan	4.5	3.1	11.1	8.0
Minnesota	4.4	2.8	11.9	9.1
Mississippi	6.2	4.0	12.0	8.0
Missouri	4.5	3.1	10.0	6.9
Montana	5.3	3.3	11.1	7.8
Nebraska	4.1	3.0	9.4	6.4
Nevada	5.3	2.8	11.0	8.2
New Hampshire	5.2	2.5	12.7	10.2
New Jersey	5.1	2.7	12.5	9.8
New Mexico	6.3	3.0	13.4	10.4
New York	6.6	3.4	14.5	11.1
North Carolina	5.3	3.0	11.9	8.9
North Dakota	4.8	3.4	9.5	6.1
Ohio	3.9	2.6	8.8	6.2
Oklahoma	5.3	3.5	10.5	7.0
Oregon	4.1	2.7	8.5	5.8
Pennsylvania	5.5	3.3	13.3	10.0
Rhode Island	5.3	2.9	12.3	9.4
South Carolina	6.5	3.1	14.8	11.7
South Dakota	4.5	3.1	9.3	6.2
Tennessee	5.1	3.0	11.0	8.0
Texas	6.0	3.3	12.8	9.5
Utah	3.7	2.1	8.5	6.4
Vermont	6.2	2.8	13.4	10.6
Virginia	4.2	2.5	10.2	7.7
Washington	4.6	2.8	9.8	7.0
West Virginia	4.8	3.4	9.3	5.9
Wisconsin	5.0	3.0	14.8	11.8
Wyoming	5.4	2.5	10.6	8.1

ESCAP MEETING NO. 31 - 02/01/01

MINUTES

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 31

February 1, 2001

Prepared by: Annette Quinlan

The thirty-first meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 1, 2001 at 10:30. The agenda for the meeting was to discuss Census operations and quality indicators measures.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines	Gregg Robinson
Tommy Wright	Roxie Jones
Donna Kostanich	Nick Birnbaum
Raj Singh	Carolee Bush
David Whitford	Kathleen Styles
Nick Alberti	Maria Urrutia
Jennifer Reichert	Sarah Brady
Jim Treat	Annette Quinlan

I. Census Operations and Quality Indicators

Jim Treat presented some operational measures of Census address listing programs. The handouts distributed at the meeting are attached. There were twenty different operations conducted for Census 2000 which provided inputs to the development of the address list. These operations are grouped into three major categories, pre-questionnaire delivery, questionnaire delivery, and post-questionnaire delivery. The percent of addresses added from each of these categories, by region, was discussed. There were no outliers or unusual patterns in the distributions of added addresses, by category. It was noted that data from 1990 are not available to use for comparison purposes because of the inherent differences in the address list development operations from 1990 to 2000.

Jennifer Reichert then presented an overview of the Quality Assurance Program that was implemented for the Census 2000 address list development operations. The three main objectives of the QA program were: (1) to prevent errors due to lack of knowledge or understanding on the part of the lister or enumerator, (2) to control coverage and content errors, and (3) to promote continuous improvements of performance. She reviewed the tools that were used in the field to meet these QA objectives.

Jim Treat continued with a discussion to provide background information on housing unit processing. There were three operations in place which removed housing units from further census processing. The first process removed housing units which were determined by two independent Census operations to be non-existent. The second operation, the deletes process, identified housing units having conflicting information from different field operations. Rules were established to determine the final status of the housing unit for census processing. The third operation identified and removed duplicate housing units through person and address matching algorithms. A table was provided showing the total number and percent of housing units removed from census processing. This information was intended for background information since these operations are not directly comparable to 1990 given the construct of the address listing operations.

Nick Alberti next presented an overview of the Primary Selection Algorithm (PSA) processing. The different types of PSA households were discussed. He then showed the Committee several different tables summarizing the data into categories such as the number of returns per census housing unit, census returns per PSA household, the number of duplicate returns for a

PSA household, and the PSA effect on population counts. These results are in line with what had been expected from the PSA operation.

Lastly, the census imputation rates for age, sex, race, Hispanic Origin, and tenure by type of response (self-administered or enumerator-administered) were presented and discussed. A few notable items from these results are: the imputation rates for the Midwest self-administered responses are consistently lower than the rates for the other regions for all categories; again raising concerns regarding the synthetic assumption. Age has the largest national imputation rate of all the characteristics, but this may be attributable to the enumerator questionnaire having both age and date of birth in the same question. Enumerators may have only asked for date of birth, in an effort to speed up the interview and possibly assuming that age could be computed later. A final note is that the imputation rates by state are mostly consistent across the five variables. That is, if a state has a low imputation rate for age it will generally have low imputation rates for the other four variables too.

Data were presented for the followup operations and Mail Response rates, but the Committee noted some discrepancies and referred the material for additional review. These data will be discussed at a future meeting.

II. Next Meeting

The next meeting scheduled for Friday February 2, 2001 will discuss A.C.E. interviewing quality assurance and After Followup Matching results.

ESCAP MEETING NO. 32 - 02/02/01

AGENDA

Kathleen P Porter
02/02/2001 08:49 AM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc: Tamara S Adams/DSSD/HQ/BOC@BOC, Danny R Childers/DSSD/HQ/BOC@BOC

Subject: Agenda for 2/2 ESCAP

The agenda for the February 2 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. A.C.E. Interviewing QA Results - Tammy Adams

2. A.C.E. After Follow up Matching Results - Danny Childers

ESCAP MEETING NO. 32 - 02/02/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Figure 1. 90% Confidence Intervals for 2000 A.C.E. Undercount Rates for Major Demographic Groups

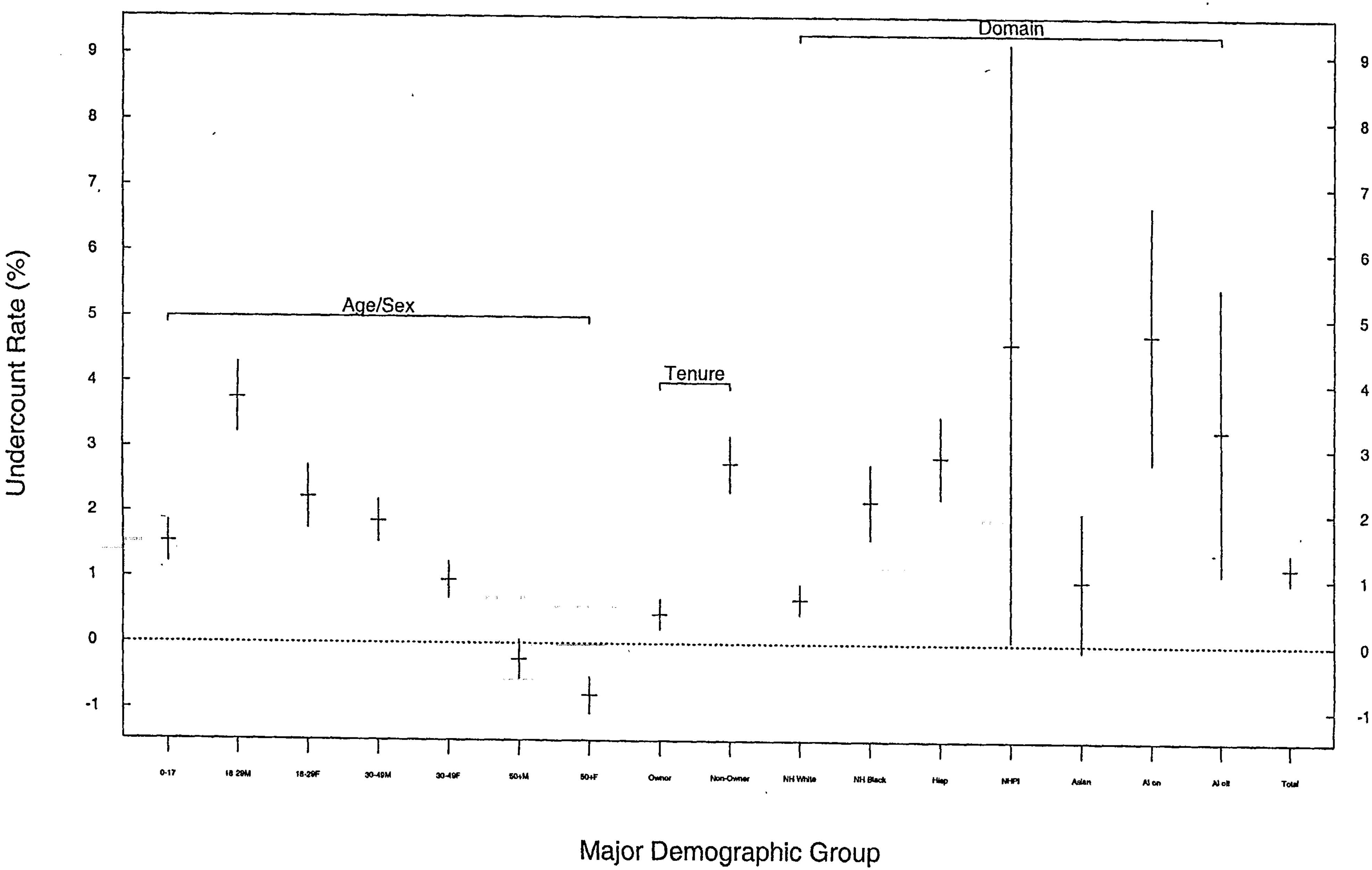


Figure 2. 90% Confidence Intervals for 2000 A.C.E. Undercount Rates Domain by Tenure

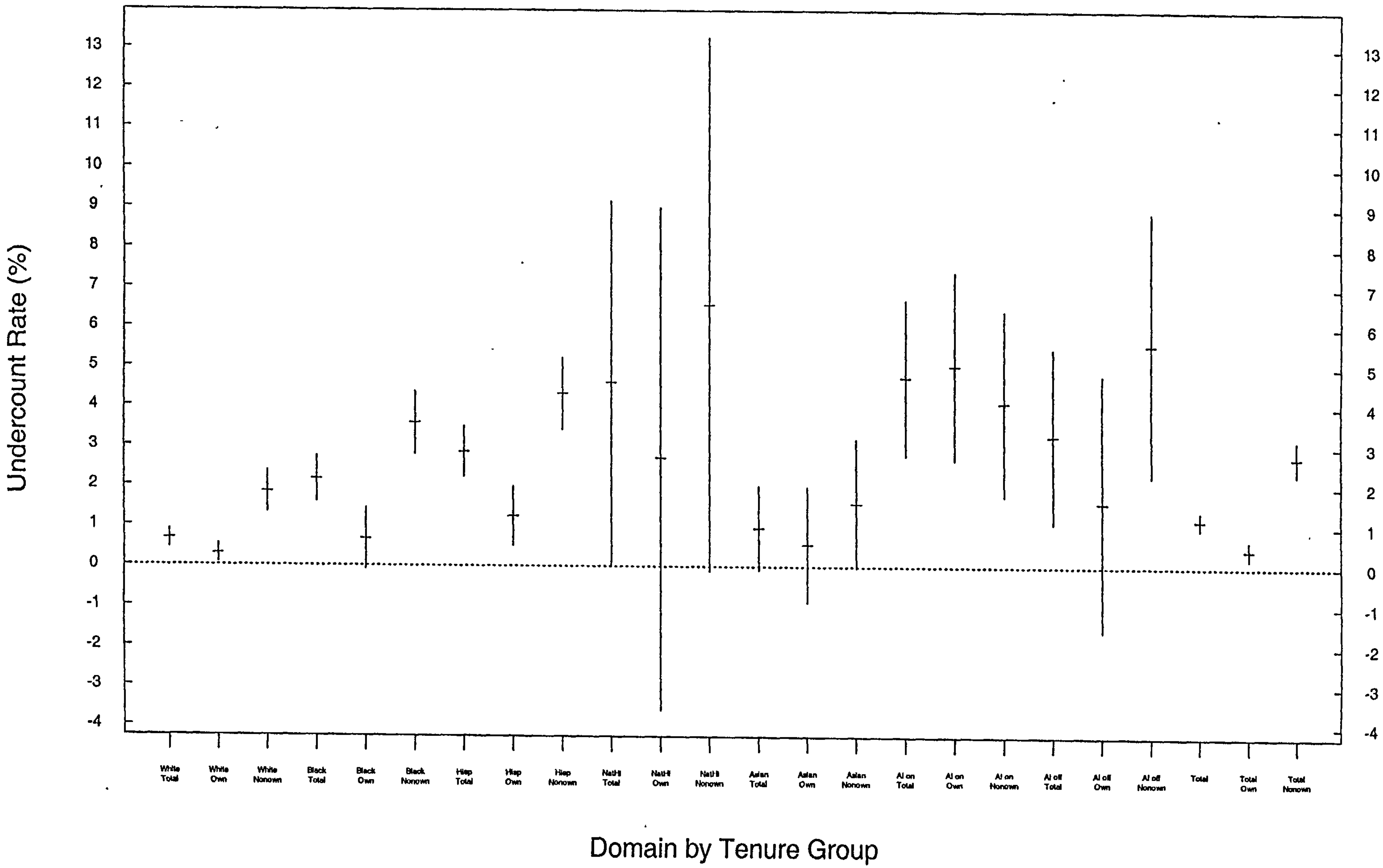


Figure 3. 90% Confidence Intervals for 2000 A.C.E. Undercount Rates
Tenure by Age/Sex

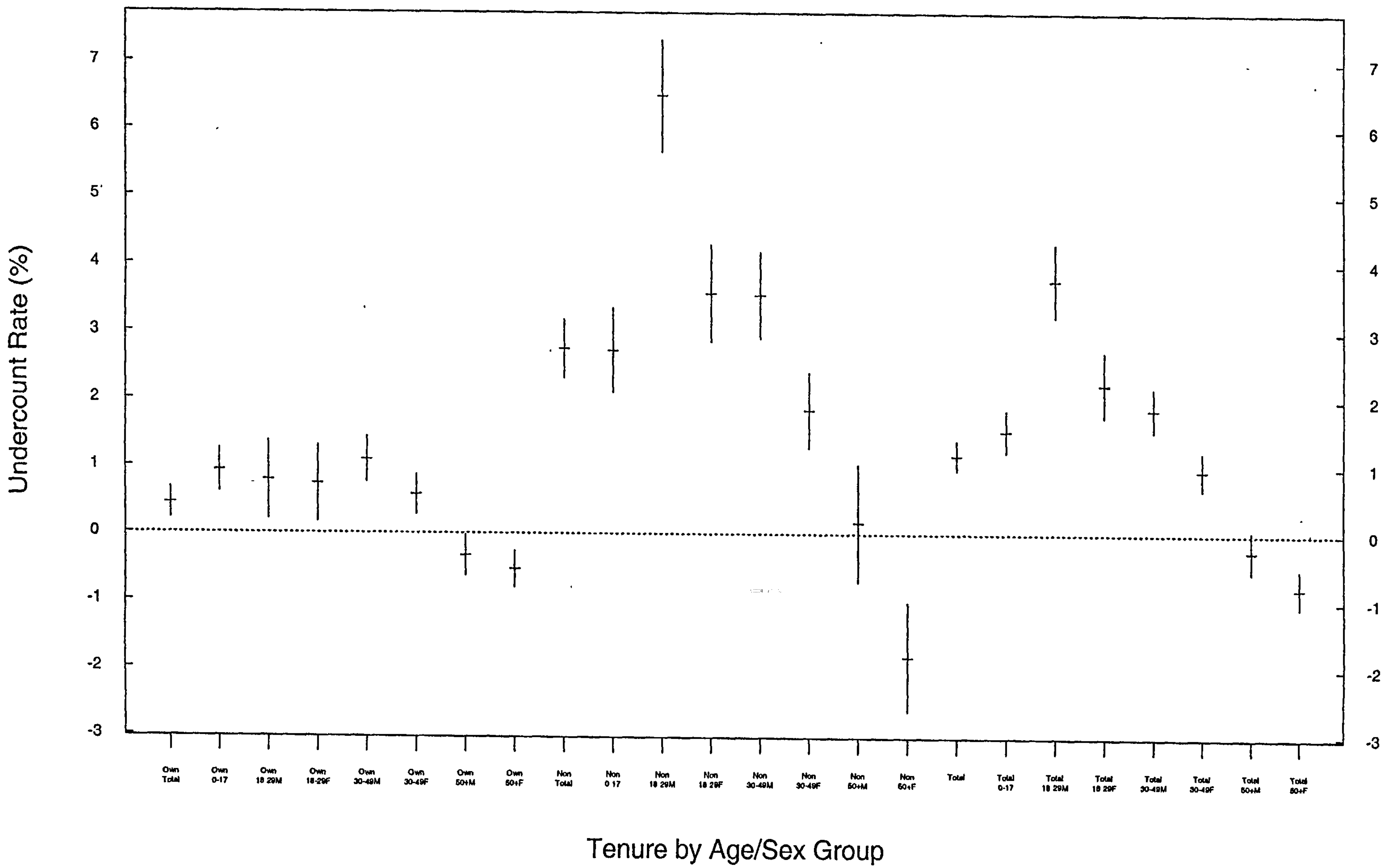


Figure 4. 90% Confidence Intervals for 2000 A.C.E. Undercount Rates
Domain by Age/Sex

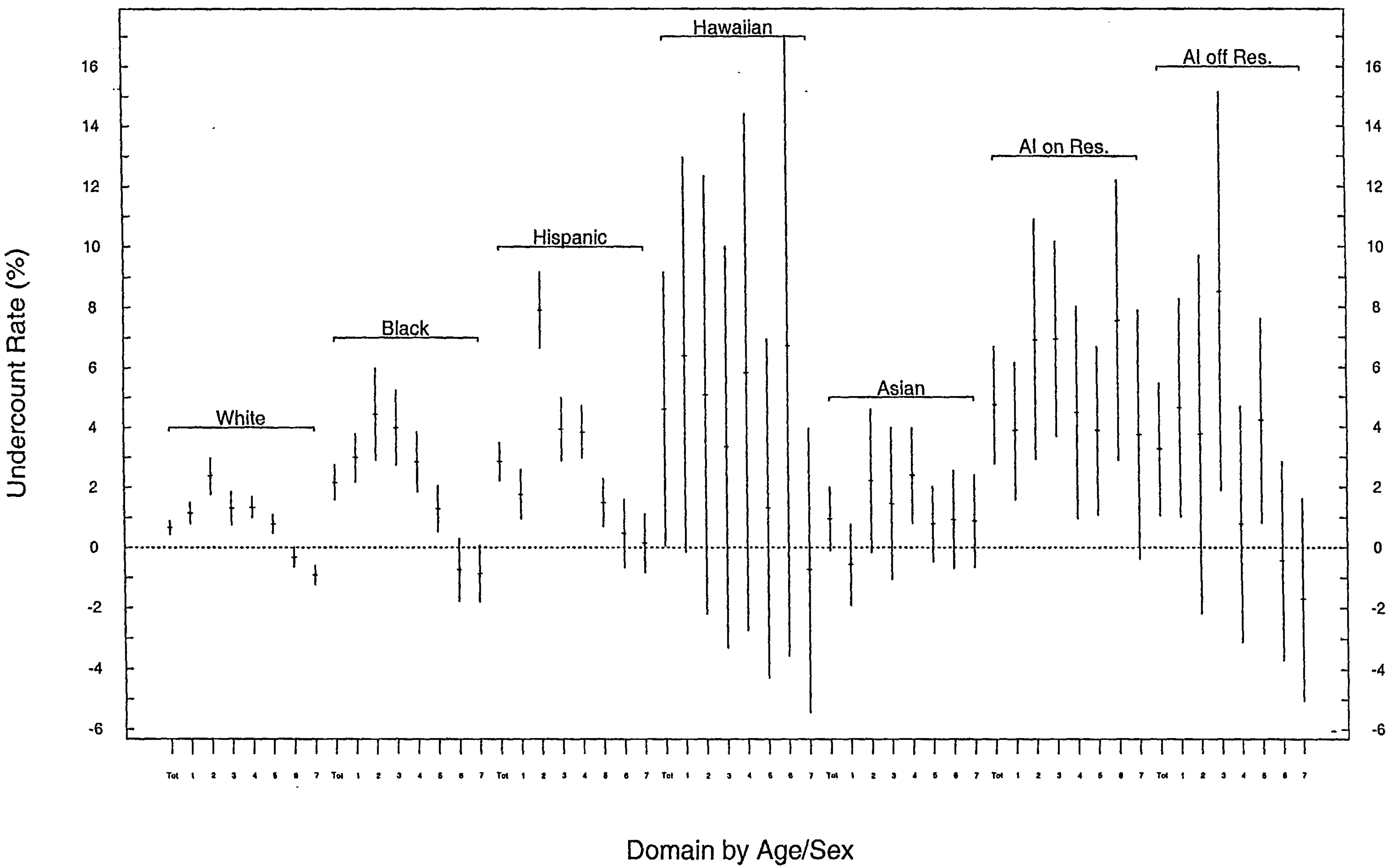
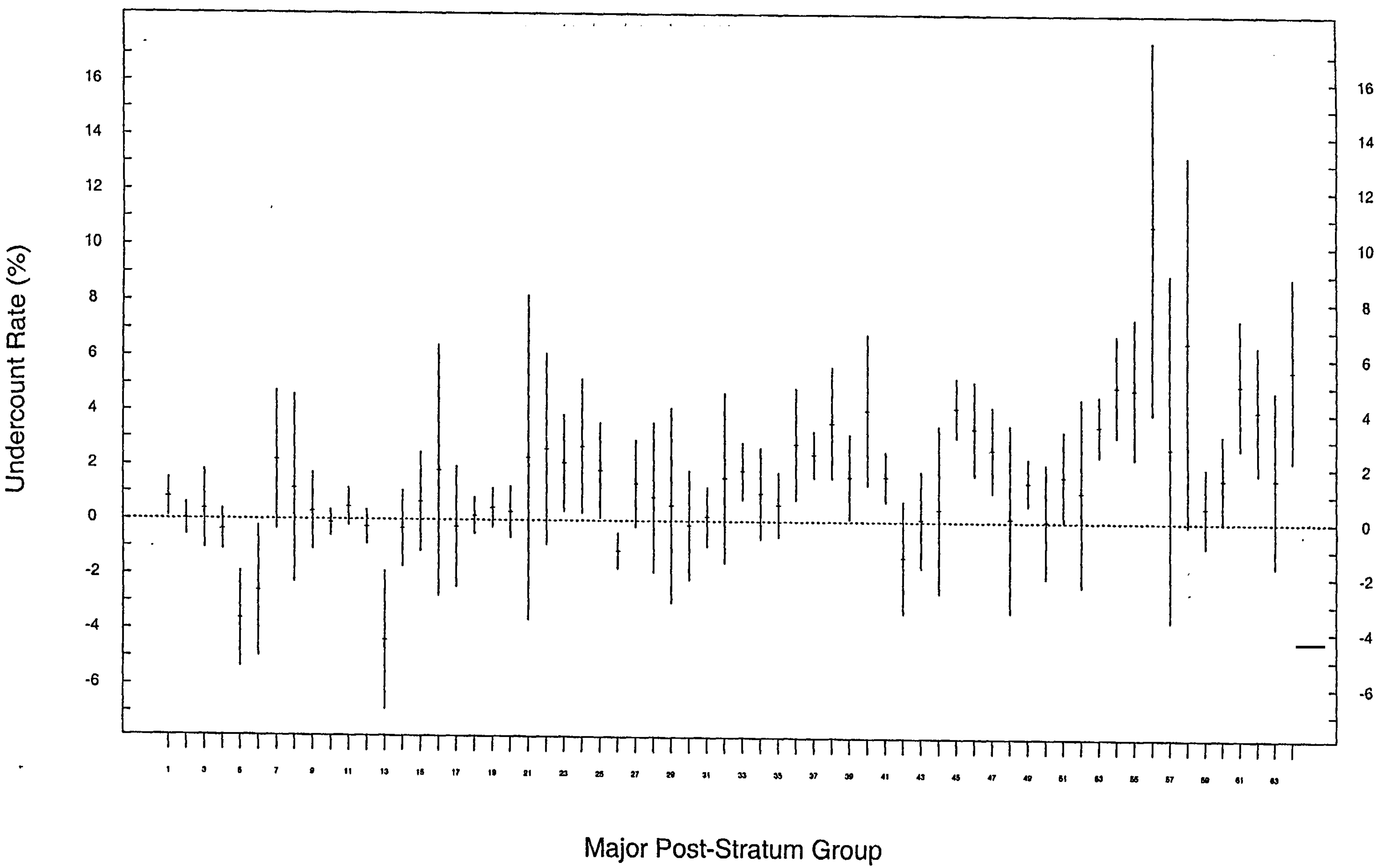


Figure 5. 90% Confidence Intervals for 2000 A.C.E. Undercount Rates
64 Major Post-Stratum Groups



Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Quality Assurance of the A.C.E. Person Interview Operation

Fast Facts:

- <u>Mode</u> — Conducted by telephone or personal visit
- <u>Interview Structure</u> — Respondents were led through a series of questions to determine if the original respondent was interviewed. If the respondent was not contacted, a full person interview is obtained from the current respondent
- <u>Replacement needed?</u>
  –The responses to questions in the QA instrument lead the QA interviewer to take a replacement interview, if necessary
  –The QA instrument decides whether a replacement interview will be used in further processing based on the QA interviewer's assessment
  –The QA supervisor in the ACERO reviews the cases that had a QA replacement interview to determine if it had discrepant results, concentrating on individual PI interviewers
- <u>Discrepant Results</u>
  –Do not include honest interviewer or respondent mistakes
  –Do include cases that failed the QA check
- <u>Discrepant?</u>
  –All of the interviewer's completed cases are assigned to QA to undergo a QA interview
  –Unfinished cases are reassigned
  –None of that interviewer's work is used in processing, unless it is determined to be "clean"

QA Sample–34,895 cases in QA, out of 300,913 total cases interviewed

- <u>Preselected:</u> — Random 5% sample chosen before interviewing begins
  Purpose: To catch random failure of QA
- <u>Targeted:</u> — Cases selected by the QA supervisors in the ACERO's who quickly monitor the PI and QA operations using computer generated reports available on-demand.

  Three online reports:
  - Outlier Report: Contained quality indicators (e.g., # cases

late at night or early in the morning, # vacants, # cases without a respondent telephone number) that could show if an interviewer was an outlier in his/her geographic area for a particular category

- Respondent Names Report: Contained respondent names to allow supervisors to target cases with missing or suspicious names
- Not Enough Cases in QA Report: Allowed supervisors to target interviewers who do not have enough work in QA

HQ Reports and Review

- Insufficient Names Report: Generated weekly at HQ and sent to the ACERO's. Contained those cases that would be considered noninterviews during matching due to insufficient information (e.g., "Refused Smith")

- FLD HQ Review: Monitoring of the ACERO investigation into the cases that were replaced. FLD HQ would review and request additional information if necessary

Bonus Review

- During matching and followup, certain whole household nonmatches are sent back to the field for a followup interview. Any cases where the P-sample people are determined to not exist are removed from the P-sample.

- <u>Excluded from Formal QA</u>

Nonresponse Conversion (NRCO) cases are excluded from QA. We attempted to use the best interviewers. In addition, virtually all interviewers used in NRCO worked during telephoning or personal visit and had already had worked checked. Those interviewers that did not work in earlier A.C.E. operations were current survey interviewers who had been through a comparable QA process.

Assumptions

- Good quality of CAPI PI interview–Due to the data edits and automated skip patterns in the PI interview, we assume that if the correct household was contacted, that the quality of the data is high.

- PI QA catches blatant mistakes–We target those interviewers which blatantly misrepresent data. We would have to have a prohibitively large sample to identify cases in which the interviewer does not misrepresent most of his or her cases.

- No QA of the QA–The QA interview is determined to be correct.

Limitation

- Cases can have replacement interviews and not be discrepant–Honest mistakes can lead to a QA replacement case. These cases are not discrepant and are not included as such in these results. Examples include: the original interviewer inadvertently conducted the interview at the wrong housing unit because the map was difficult to read, or the respondent was elderly and could not remember the original interview but recognized the questions as the interview got underway.

Results

Table 21 Outcome of QA Cases by Method of Selection

QA Results	Randomly Preselected		Targeted	
TELEPHONE PHASE				
Pass	4,398	(99.95%)	4,622	(99.52%)
Fail	2	(0.05%)	17	(0.37%)
Undetermined	0	(0.00%)	5	(0.11%)
SUBTOTAL - - Telephone	4,400	(100%)	4,644	(100%)
PERSONAL VISIT PHASE				
Pass	10,309	(99.70%)	15,329	(98.83%)
Fail	17	(0.16%)	154	(0.99%)
Undetermined	14	(0.14%)	28	(0.18%)
SUBTOTAL - - Personal Visit	10,340	(100%)	15,511	(100%)
COMBINED TOTALS				
Pass	14,707	(99.78%)	19,951	(98.99%)
Fail	19	(0.13%)	171	(0.85%)
Undetermined	14	(0.09%)	33	(0.16%)
TOTAL	14,740		20,155	

Key Points:

- Telephone and personal visit phase QA failure rates are similar
- Targeting is more effective than random preselection

Table 22. Number of Interviewers Failing QA by Region

Region	Number of interviewers with one or more interviews failing QA	Total number of interviewers whose work was QA'd	Percentage of interviewers failing QA	Total number of interviewers[1] (including supervisors)
Boston	1	581	0.2	610
New York	5	372	1.3	398
Philadelphia	2	464	0.4	501
Detroit	8	373	2.1	396
Chicago	0	389	0	402
Kansas City	2	405	0.5	411
Seattle	2	400	0.5	425
Charlotte	0	549	0	579
Atlanta	0	346	0	363
Dallas	16	568	2.8	593
Denver	2	617	0.3	625
Los Angeles	4	419	0.9	428
TOTAL	42	5483	0.8	5731

[1]248 interviewers (less than 5 percent) did not have a QA check of their work if they worked very few cases and then quit, if most of their cases were eligible for NRCO. Additionally, these totals include supervisors who may have only done a few cases as well as experienced interviewers from other surveys brought on to help in NRCO.

1990 PES vs. 2000–A.C.E. QA

Key Point

- Given the different QA designs, these numbers are not directly comparable. We can only conclude that order of magnitude differences may indicate a difference between the two failure rates.

- There are no major differences between the failure rates from 1990 PI QA and the failure rates from 2000 PI QA.

	1990[1]	2000
QA Results	Any whole household fabrication	Only discrepant results (excludes interviewer or respondent honest error)
Mode of PI Interview	Paper and pencil	CAPI
QA Interview	respondent check and dependent roster check	respondent check
# Failures	420	19 (randomly selected) 171 (targeted)
# Cases in QA	56,000	34,895
Failure Rate	0.75%	0.13% (randomly selected) 0.85% (targeted)

[1] Tremblay, 1991

Conclusions

- A.C.E. QA detected errors. Of all households subject to a QA reinterview (34,895), 979 had replacement interviews; 190 failed the QA check.

- Targeting for potential discrepant results was successful. The failure rate for targeted cases was 0.85%, compared with 0.13% in randomly preselected cases.

- The quality of the person interview cases not checked by QA is high. We had 286,173 of PI cases that were not checked in sample QA; it can be assumed that the remaining error rate would be similar to the preselected cases.

References

Tremblay, Antoinette. 1990 Decennial Census Preliminary Research and Evaluation Memorandum No. 57. Final Report for 1990 PES Evaluation Project P5: Analysis of PES P-Sample Fabrications for PES Quality Control Data

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

February 2, 2001

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Danny R. Childers
Roxanne Feldpausch
Xijian Liu
Decennial Statistical Studies Division

Subject: Accuracy and Coverage Evaluation: Results from After Follow-up Person Matching

Attached are results from the after follow-up person matching. The matching results in the first three tables are unweighted for P-sample and the E-sample. The remainder of the tables contain data that is weighted with the sampling weight.

Listing of Tables

A-9 Percent of the Weighted E-Sample by Percent Multi-Unit 34

After Follow-up Person Match Results

The final P-sample results are in Tables 1 and 2. The P-sample people have been classified as matched, not matched, unresolved match status, and removed in Table 1 and also tabulated as resident, nonresident, and unresolved residence status in Table 2.

The P-sample match status is defined as

- matched
- not matched
- unresolved match status
- removed from the P-sample

Matched - The P-sample was found in the cluster or in the surrounding block in either a housing unit or in group quarters.

Not matched - The P-sample person was not found in the search area. If the nonmatch was sent to follow-up, the person was confirmed to be a resident of the cluster on census day. If the nonmatch was not sent for a follow-up interview, a household member identified the person as a resident of the housing unit during the original A.C.E. interview.

Unresolved match status - The match status is unresolved for possible matches with unsuccessful follow-up interviews and for P-sample people with insufficient information for matching and follow-up.

Removed from the P-sample - People are removed from the P-sample when they are fictitious, duplicates, geocoding errors, or not residents of the housing unit on census day.

The P-sample residence status is defined as

- resident
- nonresident
- unresolved residence status

Resident - The P-sample matched or not matched person is a resident of the housing unit on census day.

Nonresident - P-sample people are nonresidents of the cluster when they are fictitious, duplicates, geocoding errors, or should not have been included as a resident of the housing unit on census day. Nonresidents are removed from the P-sample.

Unresolved Residence Status - A matched or not matched P-sample person has unresolved residence status when the follow-up interview did not successfully determine the person's residence on census day. The residence status of the possible match is unresolved when the follow-up interview was not successful. The residence status is also imputed when the P-sample person had insufficient information for matching.

The final E-sample results are in Table 3. The E-sample people have been classified as correctly or erroneously enumerated and enumeration status of unresolved. These are the unweighted match results that go to imputation and estimation.

The E-sample enumeration status is defined as

- correctly enumerated
- erroneously enumerated
- unresolved enumeration status

Correctly Enumerated - E-sample people are correctly enumerated when they are matched to the P-sample or when they have been followed up and they should have been enumerated in this cluster.

Erroneously Enumerated - E-sample people are erroneously enumerated when they have another residence where they should be counted on census day, are fictitious, are duplicated, lived in a housing unit that was a geocoding error, or have insufficient information for matching and follow-up.

Unresolved Enumeration Status - E-sample people have unresolved enumeration status when the follow-up interview was unsuccessful. The E-sample person may have been followed up to obtain information about the E-sample nonmatch, possible match, matched person with unresolved residence status, or geographic work to obtain the location of the housing unit.

Table 1: National P-Sample Match Status After Follow-up

P-sample Match Status	Unweighted People	Percent
Matched	578,695	86.5
Not Matched	69,551	10.4
Unresolved	7,829	1.2
Removed	12,752	1.9
Total	668,827	100.0

Table 2: National P-Sample Residence Status After Follow-up

P-sample Residence Status	Unweighted People	Percent
Resident	640,541	95.8
Nonresident	12,752	1.9
Unresolved	15,534	2.3
Total	668,827	100.0

Table 3: National E-Sample Matching After Follow-up

E-sample Enumeration Status	Unweighted People	Percent
Correctly Enumerated	652,393	91.5
Erroneously Enumerated	39,841	5.6
Unresolved	20,666	2.9
Total	712,900	100.0

The final P-sample is in Table 4 after the people are removed from the P-sample.

Table 4: Match Status for P-Sample After Follow-up

P-sample Match Status	Unweighted People	Percent
Matched	578,695	88.2
Not Matched	69,551	10.6
Unresolved	7,829	1.2
Total	656,075	100.0

Table 5 contains the net undercount from the PES in 1990 by race and ethnic origin.

Table 5: 1990 Percent Undercount from Dual System Estimation by Race and Ethnic Origin

Race and Ethnic Origin	Percent Undercount
Non-Hispanic White and Other	0.7
Black	4.6
Hispanic	5.0
Asian and Pacific Islander	2.4
Reservation Indian	12.2
Total	1.6

The percent P-sample not matched and E-sample erroneous enumeration is contained in the next set of tables. All data is weighted. The percent P-sample not matched is one hundred times the nonmatch rate.

$$\text{Nonmatch Rate} = \frac{\text{Not Matched}}{\text{Matched + Not Matched}}$$

The percent E-sample erroneous enumeration is one hundred times the erroneous enumeration rate.

$$\text{Erroneous Enumeration rate} = \frac{\text{Erroneous Enumeration}}{\text{Correct Enumeration + Erroneous Enumeration}}$$

A simple imputation is done for the unresolved people. The ratio is estimated by 1.0 minus the match rate over the correct enumeration rate, which is a coverage rate of the data defined HCUF Prime.

This is not the same as the ratio and coverage factor calculated from the dual system estimate of the population that will be calculated after all processing is completed using

- nonintervew adjustment
- TES sampling weights
- weight trimming
- a more sophisticated imputation for unresolved cases
- whole person imputations in the census

This is a measure of what we can expect after all of the estimation has been completed. Standard errors have been calculated using VPLX. Notice that blank is included for each characteristic because the missing data work had not been completed when this data was prepared.

Table 6: Percent Not Matched, Erroneously Enumerated, and Ratio After Follow-up by Census Region

Census Region	P-sample Percent Not Matched		E-sample Percent Erroneous Enumeration		Ratio	
	Percent	Standard Error	Percent	Standard Error	Percent	Standard Error
Northeast	9.4	0.2	6.0	0.2	3.6	0.3
Midwest	7.2	0.2	4.7	0.2	2.6	0.2
South	10.1	0.2	5.8	0.2	4.5	0.2
West	9.8	0.2	6.1	0.6	4.0	0.6
Total	9.2	0.1	5.7	0.2	3.8	0.2

Table 7: Percent Not Matched, Erroneously Enumerated, and Ratio After Follow-up by Census Regional Office

Census Regional Office	P-sample Percent Not Matched		E-sample Percent Erroneous Enumeration		Ratio	
	Percent	Standard Error	Percent	Standard Error	Percent	Standard Error
Boston	8.4	0.4	5.1	0.3	3.4	0.4
New York	11.5	0.5	7.5	0.4	4.4	0.5
Philadelphia	9.5	0.5	5.5	0.3	4.3	0.5
Detroit	7.0	0.3	4.6	0.3	2.5	0.4
Chicago	8.5	0.3	5.1	0.3	3.6	0.4
Kansas City	6.4	0.3	4.4	0.2	2.0	0.3
Seattle	9.7	0.4	6.0	0.4	3.9	0.5
Charlotte	9.1	0.3	5.4	0.3	3.9	0.4
Atlanta	10.5	0.4	6.4	0.3	4.4	0.4
Dallas	10.8	0.4	6.0	0.4	5.1	0.5
Denver	9.6	0.5	5.5	0.3	4.3	0.5
Los Angeles	9.7	0.4	6.4	1.5	3.6	1.5
Total	9.2	0.1	5.7	0.2	3.8	0.2

Table 8: Percent Not Matched, Erroneously Enumerated, and Ratio After Follow-up by Tenure

Tenure	P-sample Percent Not Matched		E-sample Percent Erroneous Enumeration		Ratio	
	Percent	Standard Error	Percent	Standard Error	Percent	Standard Error
Owner	7.1	0.1	3.9	0.1	3.3	0.1
Renter	14.1	0.2	7.7	0.4	6.9	0.5
Blank	13.5	0.6	21.3	0.5	-9.8	1.0
Total	9.2	0.1	5.7	0.2	3.8	0.2

Table 9: Percent Not Matched, Erroneously Enumerated, and Ratio After Follow-up by Type of Enumeration Area

Type of Enumeration Area	P-sample Percent Not Matched		E-sample Percent Erroneous Enumeration		Ratio	
	Percent	Standard Error	Percent	Standard Error	Percent	Standard Error
Mail Out/Mail Back	9.1	0.1	5.8	0.2	3.5	0.2
Update/Leave	9.5	0.3	5.1	0.2	4.7	0.3
List/Enumerate	18.2	3.5	6.2	2.0	12.9	5.1
Rural Update/Enumerate	18.1	1.3	5.4	0.7	13.4	1.3
Urban Update/Leave	7.4	1.8	4.6	1.3	3.0	1.5
Urban Update/Enumerate	8.7	3.6	1.9	1.2	6.9	3.1
Adds to Address List	13.4	4.1	4.6	1.6	9.3	3.8
Total	9.2	0.1	5.7	0.2	3.8	0.2

Table 10: Percent Not Matched, Erroneously Enumerated, and Ratio After Follow-up by Age and Sex

Age and Sex		P-sample Percent Not Matched		E-sample Percent Erroneous Enumeration		Ratio	
		Percent	Standard Error	Percent	Standard Error	Percent	Standard Error
Under 18	Male and Female	9.7	0.1	3.9	0.1	6.1	0.2
18 to 29	Male	14.4	0.2	6.8	0.4	8.1	0.5
18 to 29	Female	12.1	0.2	6.3	0.6	6.1	0.6
30 to 49	Male	9.5	0.1	4.4	0.2	5.4	0.2
30 to 49	Female	7.8	0.1	3.8	0.2	4.1	0.2
50 and over	Male	7.1	0.1	4.1	0.1	3.1	0.2
50 and over	Female	6.4	0.1	4.0	0.1	2.5	0.1
Blank	Male	14.5	0.7	41.5	0.8	-46.1	2.3
Blank	Female	12.6	0.6	41.6	0.9	-49.6	2.4
18 to 29	Blank	14.7	1.1	59.9	1.4	-112.9	8.0
30 to 49	Blank	12.3	0.9	10.1	0.9	2.4	1.4
50 and over	Blank	10.1	0.8	12.6	1.1	-2.8	1.6
Total		9.2	0.1	5.7	0.2	3.8	0.2

Table 11: Percent Not Matched, Erroneously Enumerated, and Ratio After Follow-up by Race Domain

Race Domain	P-sample Percent Not Matched		E-sample Percent Erroneous Enumeration		Ratio	
	Percent	Standard Error	Percent	Standard Error	Percent	Standard Error
Blank	16.4	1.2	7.8	0.6	9.4	1.4
American Indians on Reservations	21.2	1.4	6.3	1.0	15.9	1.6
American Indians Not on Reservations	13.9	1.2	6.1	0.6	8.3	1.3
Hispanic	13.2	0.3	8.7	0.3	4.9	0.4
Black	13.9	0.3	7.4	0.2	7.0	0.3
Pacific Islander	14.6	1.9	5.7	1.1	9.5	2.1
Asian	9.9	0.5	7.0	1.2	3.2	1.4
White	7.4	0.1	4.6	0.1	3.0	0.1
Other Races	11.8	0.7	7.6	0.7	4.5	1.0
Multiple Races	10.1	0.4	4.7	0.5	5.6	0.6
Total	9.2	0.1	5.7	0.2	3.8	0.2

Table 12: 1990 Percent Net Undercount by Race and Tenure

Race Domain	Net Undercount		
	Total	Owner	Renter
Non-Hispanic White and Other	0.7	-0.3	3.1
Black	4.6	2.3	6.5
Hispanic	5.0	1.8	7.4
Asian and Pacific Islander	2.4	-1.4	7.0
Reservation Indian	12.2	N/A	N/A

Table 13: 2000 Percent Estimated Ratio by Race Domain and Tenure

Race Domain	Ratio		
	Total	Owner	Renter
Blank	9.4	6.4	16.1
American Indians on Reservations	15.9	16.2	14.7
American Indians Not on Reservations	8.3	6.7	11.5
Hispanic	4.9	5.0	8.0
Black	7.0	6.1	9.8
Pacific Islander	9.5	7.2	13.9
Asian	3.2	3.7	3.3
White	3.0	2.8	5.5
Other Races	4.5	4.5	7.0
Multiple Races	5.6	3.9	8.7
Total	3.8	3.3	6.9

Table 14: Percent Not Matched, Erroneously Enumerated, and Ratio After Follow-up by Mail Return Rate						
Mail Return Rate	P-sample Percent Not Matched		E-sample Percent Erroneous Enumeration		Ratio	
	Percent	Standard Error	Percent	Standard Error	Percent	Standard Error
Greater than 25 percent	7.7	0.1	5.0	0.2	2.8	0.2
Less than 25 percent	15.7	0.3	8.8	0.2	7.6	0.4
List/Enumerate, Rural Update Enumerate, Urban Update/Enumerate	17.6	1.3	5.6	0.9	12.7	1.9
No Occupied Housing Units or Incomplete Address	37.7	5.8	0.0	--	37.7	5.8
Total	9.2	0.1	5.7	0.2	3.8	0.2

We do not have housing variables in the census to indicate multi-units and mobile homes. We created variables by classifying the clusters based on the housing unit variables from the ACE listing. The next two tables contain percent not matched and erroneous enumeration for clusters different amounts of multi-units and mobile homes.

Table 15: Percent Not Matched, Erroneously Enumerated, and Ratio After Follow-up by Percent Mobile Home

Percent Mobile Home	P-sample Percent Not Matched		E-sample Percent Erroneous Enumeration		Ratio	
	Percent	Standard Error	Percent	Standard Error	Percent	Standard Error
None	8.8	0.1	5.7	0.2	3.4	0.2
10 Percent or less	9.0	0.4	5.3	0.3	3.9	0.4
11 to 50 percent	10.7	0.4	5.5	0.3	5.5	0.4
Greater than 50 percent	13.8	0.9	7.2	0.4	7.1	1.0
Total	9.2	0.1	5.7	0.2	3.8	0.2

Table 16: Percent Not Matched, Erroneously Enumerated, and Ratio After Follow-up by Percent Multi-Unit

Percent Multi-Unit	P-sample Percent Not Matched		E-sample Percent Erroneous Enumeration		Ratio	
	Percent	Standard Error	Percent	Standard Error	Percent	Standard Error
None	8.2	0.1	4.9	0.1	3.5	0.2
10 Percent or less	7.8	0.3	4.4	0.2	3.6	0.3
11 to 50 percent	9.4	0.3	6.0	0.2	3.6	0.3
Greater than 50 percent	12.9	0.3	8.2	0.7	5.1	0.7
Total	9.2	0.1	5.7	0.2	3.8	0.2

<table>
<caption>Table 17: Ratio Estimates by Simulated Post-Strata Variables</caption>
<tr><th rowspan="2">Race Domain</th><th rowspan="2">Tenure</th><th rowspan="2">TEA</th><th colspan="4">High Return Rate</th><th colspan="4">Low Return Rate</th></tr>
<tr><th>Northeast</th><th>Midwest</th><th>South</th><th>West</th><th>Northeast</th><th>Midwest</th><th>South</th><th>West</th></tr>
<tr><td rowspan="4">White and Other</td><td rowspan="2">Owner</td><td>MO/MB</td><td>2.0</td><td>1.5</td><td>2.5</td><td>2.2</td><td>4.8</td><td>7.8</td><td>5.9</td><td>9.1</td></tr>
<tr><td>Other</td><td>4.6</td><td>2.5</td><td>3.8</td><td>5.1</td><td>5.4</td><td>4.8</td><td>7.4</td><td>16.7</td></tr>
<tr><td rowspan="2">Renter</td><td>MO/MB</td><td colspan="4">4.0</td><td colspan="4">8.6</td></tr>
<tr><td>Other</td><td colspan="4">7.2</td><td colspan="4">12.3</td></tr>
<tr><td rowspan="4">Black</td><td rowspan="2">Owner</td><td>MO/MB</td><td colspan="4">5.9</td><td colspan="4">6.6</td></tr>
<tr><td>Other</td><td colspan="4">5.3</td><td colspan="4">7.8</td></tr>
<tr><td rowspan="2">Renter</td><td>MO/MB</td><td colspan="4">7.8</td><td colspan="4">11.2</td></tr>
<tr><td>Other</td><td colspan="4">12.4</td><td colspan="4">13.3</td></tr>
<tr><td rowspan="4">Hispanic</td><td rowspan="2">Owner</td><td>MO/MB</td><td colspan="4">4.5</td><td colspan="4">6.6</td></tr>
<tr><td>Other</td><td colspan="4">5.8</td><td colspan="4">7.6</td></tr>
<tr><td rowspan="2">Renter</td><td>MO/MB</td><td colspan="4">5.8</td><td colspan="4">10.7</td></tr>
<tr><td>Other</td><td colspan="4">15.4</td><td colspan="4">9.4</td></tr>
<tr><td rowspan="2">Pacific Islander</td><td colspan="2">Owner</td><td colspan="8">7.2</td></tr>
<tr><td colspan="2">Renter</td><td colspan="8">13.9</td></tr>
<tr><td>Race Domain</td><td colspan="2">Tenure</td><td colspan="8">All Regions and Return Rates</td></tr>
</table>

Asian	Owner	3.7
	Renter	3.3
American Indians on Reservations	Owner	16.2
	Renter	14.7
American Indians not on Reservations	Owner	6.7
	Renter	11.5
Multiple Races	Owner	3.9
	Renter	8.7
Domain or Tenure Blank		-6.4

Table 18: Percent Not Matched After Follow-up by ACE Mover Status

Mover Status	P-sample Percent Not Matched	
	Percent	Standard Error
Nonmover	8.5	0.1
Outmover	22.9	0.5
Unresolved Mover Status	26.2	1.0
Total	9.2	0.1

Table 19: Percent Not Matched After Follow-up by ACE Interview Mode

Interview Mode	P-sample Percent Not Matched	
	Percent	Standard Error
Telephone	1.5	0.0
Personal Visit	14.0	0.2
Quality Assurance Replacement	13.4	1.7
Total	9.2	0.1

Table 20: Percent Not Matched After Follow-up by ACE Respondent Type		
Respondent Type	P-sample Percent Not Matched	
	Percent	Standard Error
Household Member	8.7	0.1
Proxy Respondent	21.7	0.5
Total	9.2	0.1

Erroneously enumerated - The categories are people with insufficient information for matching and follow-up, duplicates, fictitious, geocoding errors, and people who should have been enumerated at another residence on census day.

- The E-sample people with insufficient information for matching and follow-up are ones who are data defined, but do not contain full name and at least two characteristics.
- The E-sample people enumerated more than once are coded as duplicates.
- The fictitious people are ones where we found notes on the census image identifying the person as not a real person such as a dog or other pet or they were identified as not existing in this cluster during the follow-up interview. Three respondents who never heard of the person were required in order to code a person as fictitious.
- Census people in housing units identified as geocoding errors during the housing unit follow-up are coded as erroneously enumerated because of geocoding error.
- The E-sample person should have been counted at another residence on census day.

The next table contains the final weighted and imputed data after person follow-up in 1990 for the erroneous enumerations.

Table 21: 1990 Erroneous Enumerations Final Weighted Numbers		
E-sample Erroneous Enumeration Code	Percent of Erroneous Enumerations	Percent of E-sample
Insufficient Information	20.8	1.2
Duplicate	28.2	1.6
Fictitious	2.6	0.2
Geocoding Error	6.0	0.3
Other Residence	38.0	2.2
Unresolved	4.5	0.3
Total	100.0	5.8

The percentages of each type of erroneous enumeration in these tables are based on the E-sample people with a resolved enumeration status. Note that the percentage of each type of erroneous enumeration is one hundred times the rate of each type of erroneous enumeration. The percent duplicate includes the duplication between E-sample and census people not in sample after subsampling large clusters.

$$\text{Rate of Type of Erroneous Enumeration} = \frac{\text{Type of Erroneous Enumeration}}{\text{Correct Enumeration + Erroneous Enumeration}}$$

The remaining tables contain the type of erroneous enumeration as a percent of the total E-sample resolved cases by different variables.

Table 22: Percent of E-sample for Type of Erroneous Enumeration After Follow-up for Simulated Post-Strata Variables

Race Domain	Tenure	TEA	Mail Return Rate	Region	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
White and Other	Owner	MO/MB	High	Northeast	0.5	0.4	0.1	1.4	0.9
				Midwest	0.6	0.2	0.1	1.3	0.6
				South	0.7	0.2	0.1	1.2	1.0
				West	1.0	0.3	0.0	0.8	0.9
			Low	Northeast	1.0	1.8	0.7	2.6	1.2
				Midwest	1.0	1.3	0.1	0.5	0.8
				South	1.0	1.2	0.5	2.2	1.2
				West	1.2	0.7	0.2	1.4	1.2
		Other TEA	High	Northeast	0.4	0.5	0.0	0.5	1.1
				Midwest	0.4	0.4	0.0	0.4	1.1
				South	0.5	0.6	0.1	0.5	1.3
				West	0.7	0.3	0.0	0.5	1.8
			Low	Northeast	0.8	2.9	0.0	0.6	1.4
				Midwest	0.4	1.3	0.0	0.9	1.4
				South	0.6	1.7	0.1	1.4	1.4
				West	1.7	1.7	0.5	0.5	2.6

Race Domain Tenure TEA Mail Region Insufficient Duplicate Fictitious Geocoding Other

Race Domain	Tenure	TEA	Mail Return Rate	Region	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
White and Other	Renter	MO/MB	High		1.9	0.8	0.4	1.4	1.1
			Low		2.4	1.8	0.7	1.3	1.1
		Other TEA	High		1.1	0.8	0.1	0.5	2.0
			Low		1.7	2.4	0.8	0.6	1.3
Black	Owner	MO/MB	High		1.2	0.4	0.3	1.3	1.0
			Low		2.1	2.8	0.6	0.8	1.1
		Other TEA	High		1.0	1.0	0.1	0.2	2.2
			Low		0.3	2.3	0.1	0.9	0.7
	Renter	MO/MB	High		2.7	0.7	1.1	0.8	1.1
			Low		3.0	1.7	1.1	0.6	1.0
		Other TEA	High		1.0	1.1	1.2	0.2	0.8
			Low		1.3	1.2	0.4	1.7	1.7
Hispanic	Owner	MO/MB	High		2.1	0.5	0.1	0.8	0.8
			Low		1.8	1.6	0.9	1.9	1.1
		Other TEA	High		2.5	0.9	0.1	0.6	1.2
			Low		2.3	2.3	0.1	1.1	1.8
	Renter	MO/MB	High		3.2	1.1	0.4	1.5	0.9
			Low		3.7	1.9	1.0	0.8	1.1

		Other TEA	High	3.3	1.5	0.4	0.2	2.2
			Low	5.3	5.0	1.2	1.0	1.1
Pacific Islander	Owner			0.7	0.9	0.4	1.2	0.5
	Renter			1.1	0.9	0.6	1.8	0.3
Asian	Owner			0.8	0.7	0.2	1.4	1.1
	Renter			1.6	1.2	0.7	3.7	1.2
American Indians on Reservations	Owner			0.6	0.8	0.1	2.8	1.7
	Renter			0.5	0.3	0.1	3.1	1.6
American Indians Not on Reservations	Owner			2.3	0.4	0.1	0.9	0.7
	Renter			2.2	1.1	0.5	0.8	1.1
Multiple Races	Owner			1.0	0.6	0.0	0.8	1.0
	Renter			1.5	0.4	0.5	1.3	1.0
Domain or Tenure Blank				12.8	1.8	0.5	1.0	1.8

Table 23: Percent of E-sample for Type of Erroneous Enumeration After Follow-up by Race Domain

Race Domain	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Blank	3.5	0.9	0.3	1.0	1.3
American Indians on Reservations	0.6	0.7	0.1	2.8	1.7
American Indians Not on Reservations	2.5	0.8	0.3	0.8	0.9
Hispanic	4.1	1.2	0.4	1.0	1.0
Black	2.6	1.2	0.7	0.9	1.1
Pacific Islander	1.7	0.9	0.5	1.5	0.5
Asian	1.3	0.9	0.4	2.3	1.1
White	1.2	0.6	0.2	1.1	1.0
Other Races	2.7	1.1	0.7	0.9	1.1
Multiple Races	1.3	0.6	0.2	1.0	1.1
Total	1.8	0.8	0.3	1.1	1.0

Table 24: Percent of E-sample for Type of Erroneous Enumeration After Follow-up by Type of Enumeration Area

Type of Enumeration Area	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Mail Out/Mail Back	1.9	0.7	0.3	1.2	1.0
Update/Leave	1.4	1.1	0.1	0.6	1.5
List/Enumerate	1.8	0.3	0.0	1.7	0.7
Rural Update/Enumerate	1.8	0.4	0.1	1.2	1.6
Urban Update/Leave	2.5	0.2	0.2	0.7	0.8
Urban Update/Enumerate	0.4	0.0	0.0	0.7	0.6
Adds to Address List	1.5	1.0	0.0	1.0	1.0
Total	1.8	0.8	0.3	1.1	1.0

Table 25: Percent of E-sample for Type of Erroneous Enumeration After Follow-up by Age and Sex

Age and Sex		Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Under 18	Male and Female	0.9	0.5	0.2	1.1	0.7
18 to 29	Male	0.8	0.9	0.5	1.3	2.2
18 to 29	Female	0.9	0.8	0.4	1.4	1.9
30 to 49	Male	0.9	0.7	0.3	1.2	0.8
30 to 49	Female	0.8	0.6	0.3	1.2	0.6
50 and over	Male	0.8	0.7	0.2	1.0	1.2
50 and over	Female	0.9	0.7	0.1	0.9	1.1
Blank	Male	31.9	3.5	0.7	0.9	2.7
Blank	Female	32.0	3.7	0.7	0.9	2.6
18 to 29	Blank	54.9	1.9	0.6	0.6	1.3
30 to 49	Blank	5.3	1.6	0.6	1.2	0.7
50 and over	Blank	6.4	2.1	0.1	1.3	2.1
Total		1.8	0.8	0.3	1.1	1.0

Table 26: Percent of E-sample for Type of Erroneous Enumeration After Follow-up by Tenure

Tenure	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Owner	0.9	0.5	0.1	1.1	1.0
Renter	2.4	1.2	0.6	1.2	1.1
Blank	14.6	1.9	0.5	1.0	1.9
Total	1.8	0.8	0.3	1.1	1.0

Table 27: Percent of E-sample for Type of Erroneous Enumeration After Follow-up by Region

Region	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Northeast	1.8	1.2	0.3	1.1	1.1
Midwest	1.5	0.6	0.2	1.0	0.8
South	1.9	0.7	0.3	1.1	1.2
West	2.2	0.6	0.2	1.4	1.0
Total	1.8	0.8	0.3	1.1	1.0

Table 28: Percent of E-sample for Type of Erroneous Enumeration After Follow-up by Regional Office

Regional Office	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Boston	1.4	1.0	0.2	1.1	1.2
New York	2.4	2.1	0.6	0.8	1.0
Philadelphia	1.9	0.6	0.3	1.2	0.9
Detroit	1.6	0.4	0.2	1.1	0.9
Chicago	1.6	0.7	0.3	1.1	0.8
Kansas City	1.4	0.6	0.1	0.7	1.0
Seattle	2.6	0.7	0.2	0.9	0.9
Charlotte	1.7	0.7	0.3	1.0	1.1
Atlanta	2.2	0.7	0.2	1.1	1.6
Dallas	1.9	0.8	0.5	1.2	1.1
Denver	1.8	0.5	0.3	0.7	1.4
Los Angeles	1.9	0.5	0.2	2.2	0.7
Total	1.8	0.8	0.3	1.1	1.0

Table 29: Percent of E-sample for Type of Erroneous Enumeration After Follow-up by State

State	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Alaska	1.1	0.6	0.6	0.3	1.5
Alabama	1.8	1.3	0.2	0.9	1.0
Arkansas	1.6	0.8	0.2	0.4	1.0
Arizona	2.7	0.6	0.6	0.8	2.5
California	2.1	0.5	0.2	1.7	0.7
Colorado	2.0	0.3	0.1	0.7	0.8
Connecticut	1.2	0.8	0.1	1.4	1.1
District of Columbia	2.9	1.7	0.2	0.0	0.8
Delaware	2.3	0.9	0.1	3.5	1.5
Florida	2.2	0.5	0.2	1.1	2.0
Georgia	2.3	0.8	0.3	1.4	1.0
Hawaii	1.6	1.1	0.5	1.9	0.9
Iowa	1.3	0.7	0.0	0.7	1.1
Idaho	1.7	0.9	0.1	0.7	1.2
Illinois	1.8	0.9	0.5	0.9	0.8
Indiana	1.7	0.4	0.2	1.8	0.7
Kansas	1.5	0.7	0.3	0.9	1.0
Kentucky	1.0	0.7	0.1	1.5	1.1
Louisiana	1.7	1.4	0.4	1.4	0.9
Massachusetts	1.3	1.0	0.2	1.2	1.1
Maryland	2.2	0.5	0.5	0.7	0.8
Maine	1.1	1.3	0.0	0.3	1.5
Michigan	1.7	0.4	0.2	1.1	1.0

State	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Minnesota	1.1	0.3	0.0	0.3	0.9
Missouri	1.4	0.6	0.1	1.2	0.7
Mississippi	2.2	0.9	0.7	1.1	1.8
Montana	1.0	0.3	0.0	0.6	1.0
North Carolina	1.8	0.8	0.6	0.4	1.3
North Dakota	1.0	0.4	0.0	1.5	0.9
Nebraska	1.0	0.4	0.2	0.0	0.9
New Hampshire	1.0	0.8	0.0	1.5	1.2
New Jersey	1.6	1.1	0.3	1.1	1.0
New Mexico	1.9	0.9	0.2	1.9	0.8
Nevada	1.1	0.5	0.2	0.4	1.2
New York	2.2	1.8	0.5	0.9	1.1
Ohio	1.5	0.4	0.2	1.1	0.8
Oklahoma	1.5	0.4	0.1	0.7	1.8
Oregon	2.8	0.6	0.4	1.1	1.0
Pennsylvania	1.7	0.7	0.3	1.2	0.9
Rhode Island	2.6	1.7	0.5	1.2	0.9
South Carolina	2.2	1.0	0.1	1.5	1.6
South Dakota	0.7	0.7	0.1	0.1	1.2
Tennessee	1.7	0.7	0.1	1.0	0.9
Texas	1.9	0.7	0.5	1.2	1.0
Utah	1.9	0.6	0.1	0.3	0.9
Virginia	1.6	0.4	0.1	1.2	1.0
Vermont	1.5	1.5	0.1	0.9	1.5

State	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Washington	2.5	0.6	0.1	1.2	0.9
Wisconsin	1.2	0.6	0.0	0.8	0.8
West Virginia	1.4	0.6	0.4	0.8	0.8
Wyoming	1.0	1.1	0.0	1.4	1.5
Total	1.8	0.8	0.3	1.1	1.0

Attachment

Percentage of the Weighted E-sample for Each Variable

Table A-1: Percent of the Weighted E-Sample by Census Region

Census Region	Percent of E-Sample
Northeast	19.0
Midwest	22.8
South	35.6
West	22.6
Total	100.0

Table A-2: Percent of the Weighted E-Sample by Census Regional Office

Census Regional Office	Percent of E-Sample
Boston	7.2
New York	6.2
Philadelphia	8.1
Detroit	8.3
Chicago	8.5
Kansas City	7.9
Seattle	7.8
Charlotte	10.5
Atlanta	10.0
Dallas	9.7
Denver	7.1
Los Angeles	8.6
Total	100.0

Table A-3: Percent of the Weighted E-Sample by Tenure

Tenure	Percent of E-Sample
Owner	67.2
Renter	29.0
Blank	3.8
Total	100.0

Table A-4: Percent of the Weighted E-Sample by Type of Enumeration Area

Type of Enumeration Area	Percent of E-Sample
Mail Out/Mail Back	82.1
Update/Leave	16.8
List/Enumerate	0.2
Rural Update/Enumerate	0.5
Urban Update/Leave	0.2
Urban Update/Enumerate	0.0
Adds to Address List	0.2
Total	100.0

Table A-5: Percent of the Weighted E-Sample by Age and Sex

Age and Sex		Percent of E-Sample
Under 18	Male and Female	25.2
18 to 29	Male	7.4
18 to 29	Female	7.5
30 to 49	Male	14.7
30 to 49	Female	15.3
50 and over	Male	12.1
50 and over	Female	15.5
Blank	Male	1.3
Blank	Female	1.2
18 to 29	Blank	0.4
30 to 49	Blank	0.2
50 and over	Blank	0.2
Total		100.0

Table A-6: Percent of the Weighted E-Sample by Race Domain

Race Domain	Percent of E-Sample
Blank	0.7
American Indians on Reservations	0.2
American Indians Not on Reservations	0.6
Hispanic	15.5
Black	10.3
Pacific Islander	0.2
Asian	3.3
White	67.3
Other Races	0.6
Multiple Races	1.3
Total	100.0

Table A-7: Percent of the Weighted E-Sample by Mail Return Rate

Mail Return Rate	Percent of E-Sample
Greater than 25 percent	81.6
Less than 25 percent	17.7
List/Enumerate, Rural Update Enumerate, Urban Update/Enumerate	0.0
No Occupied Housing Units or Incomplete Address	0.7
Total	100.0

Table A-8: Percent of the Weighted E-Sample by Percent Mobile Home

Percent Mobile Home	Percent of E-Sample
None	76.9
10 Percent or less	9.3
11 to 50 percent	10.4
Greater than 50 percent	3.4
Total	100.0

Table A-9: Percent of the Weighted E-Sample by Percent Multi-Unit

Percent Multi-Unit	Percent of E-Sample
None	55.3
10 Percent or less	11.0
11 to 50 percent	14.2
Greater than 50 percent	19.5
Total	100.0

ESCAP MEETING NO. 32 - 02/02/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 32

February 2, 2001

Prepared by: Maria Urrutia and Sarah Brady

The thirty-second meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 2, 2001 at 10:30. The agenda for the meeting was to present results from the A.C.E. interviewing quality assurance and to present After Followup matching results. In addition, graphical representations of 90 percent confidence intervals for the undercount rates based on the Dual Systems Estimates were presented.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines	Tamara Adams
Tommy Wright	Nick Birnbaum
Donna Kostanich	Sarah Brady
Raj Singh	Carolee Bush
David Whitford	Annette Quinlan
Danny Childers	Kathleen Styles
Deborah Fenstermaker	Maria Urrutia
Roxie Jones	

I. Graphical Representations of 90 percent Confidence Intervals for Undercount Rates

Howard Hogan presented graphics of the 90 percent confidence intervals for the undercount rates based on the Dual System Estimates (DSEs). The DSEs were presented at the January 26, 2001 ESCAP meeting. The graphs are attached and will assist the Committee in their review of the data.

II. A.C.E. Interviewing Quality Assurance

Tamara Adams described the methodology that was implemented for the A.C.E. interviewing quality assurance (QA). There were two types of cases in the QA sample; preselected cases and targeted cases. The preselected cases were a random 5 percent sample chosen before interviewing began. The targeted cases were selected by the QA supervisors at the A.C.E. Regional Offices (ACEROs), who continuously monitored the interviewing and QA operations to identify potential outliers using computer generated reports that are readily available for review. The QA interview was conducted by either telephone or personal visit. During the QA interview, respondents were led through a series of questions to determine if the original household was interviewed. If the respondent indicated that the original respondent was not contacted, a full person interview was obtained from the current respondent. The respondent's answers and the interviewer's assessment of whether the original interview took place were used to determine whether a replacement interview would be used in further processing.

Tamara then presented the results for the QA of the A.C.E. interviewing. Based upon these results, the Committee noted that the QA followed the planned program and the results compared favorably to the 1990 PES results. Overall, the Committee concluded that the A.C.E. field interview was conducted very accurately, as measured by the QA program.

III. A.C.E. After Followup Matching Results

Danny Childers presented detailed results from the After Followup matching. The results were characterized by P-sample not matched and E-sample erroneous enumerations. Danny also presented a ratio which subtracts from 1 the proportion of P-sample matches divided by the proportion of E-sample correct enumerations. The ratio is a coverage indicator, but it is not the final coverage measurement resulting from the DSEs. Similar to the DSEs, the ratio gives an indication of coverage at higher levels, but the ratio can also measure coverage at levels such as Type of Enumeration Area, where the DSEs are more difficult to compute.

The Committee noticed that there was a regional effect on coverage, as measured by the ratio. The ratio in the Midwest was lower than the ratios for the other regions, indicating better coverage in the Midwest. The percent P-sample not matched and percent of E-sample erroneous enumerations were also lower for the Midwest. This raised concerns regarding the synthetic assumption. The Committee also noted that the ratio was very high for American Indians not on Reservations. However, the differential was not as large once the DSEs were calculated. The ratio was also presented for clusters by percent mobile home and percent multi-unit. There was modest variation in these data. John Thompson expressed pleasant surprise that there was only a modest variation in the data as the percent multi-units and percent mobile homes increased; he was expecting the difference to be larger. All data presented are preliminary results and the similar final results are documented in the DSSD Census 2000 Procedures and Operations Memorandum Series Number B-6. The handouts distributed at the meeting are attached.

Danny also presented results from the After Followup matching for types of erroneous enumeration for the E-sample. The types of erroneous enumeration were: insufficient information, duplicate, "fictitious"[1], geocoding error, and other residence. The Committee began to examine the data for the erroneous enumerations and will continue to review the data at the next ESCAP meeting.

IV. Next Meeting

The agenda for the next meeting, scheduled for February 6, 2001, is a continuation of the A.C.E. After Followup matching results and an examination of the preliminary total error model results.

[1]"Fictitious" cases are the cases that the A.C.E. could not determine as valid enumerations.

ESCAP MEETING NO. 33 - 02/06/01

AGENDA

Kathleen P Porter
02/05/2001 01:38 PM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc: Danny R Childers/DSSD/HQ/BOC@BOC, Mary Helen Mulry/DSSD/HQ/BOC@BOC

Subject: Agenda for 2/6 ESCAP

The agenda for the February 6 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. A.C.E. After Follow up Matching Results - Danny Childers

2. Preliminary Total Error Model Results - Mary Mulry

ESCAP MEETING NO. 33 - 02/06/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

OVERVIEW OF PRELIMINARY RESULTS FOR TOTAL ERROR MODEL

Mary H. Mulry

February 6, 2001

Purpose of preliminary total error analysis

- We preview the net effect of the sampling and nonsampling components of error in the A.C.E., with the exception of correlation bias and some imputation error.
- We preview confidence intervals for the A.C.E. estimates based on bias and variance estimates from the total error analysis prior to using the methodology in the loss function analysis comparing the original enumeration and the A.C.E. estimates.

Estimation Strategy

- First, we estimate component errors and their variances for groups of A.C.E. poststrata called evaluation poststrata, some with 1990 data.
- Then we derive estimates of component errors for each A.C.E. postratum based on the component errors for its evaluation poststrata.
- We use simulation methodology to assess the net effect of all the component errors combined and for use in the loss function analysis.

Preliminary Results

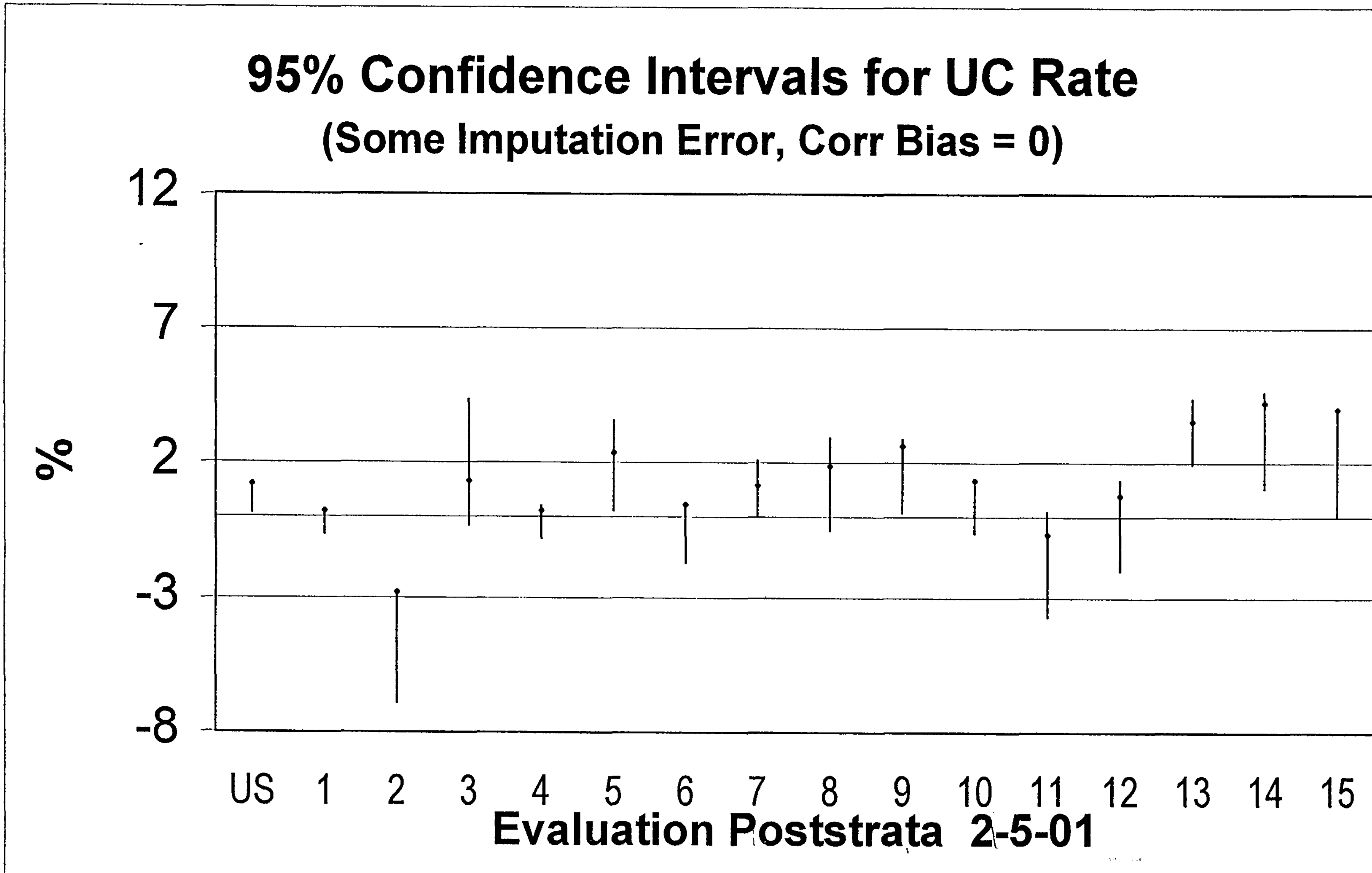
- Four of the15 Evaluation Poststrata and the US as a whole have an overcount or undercount rate that is not in its confidence interval. In 1990, two Evaluation Poststrata and the US had an undercount rate not the confidence interval when correlation bias was excluded. With correlation bias, all undercount rates were covered by their confidence intervals.
- Four of the five Evaluation Poststrata for non-owners have statistically significant undercounts.
- The measurement of the component errors is not precise enough to determine whether the undercount rates for 7 Evaluation Poststrata are different from zero. In 1990, the confidence intervals for 5 of the 10 Evaluation Poststrata covered zero when correlation bias was excluded. With correlation bias, one stopped covering zero and another started covering zero for a total of 5.
- With the inclusion of the correlation bias and the additional imputation error components, the confidence intervals will be slightly wider and shift in a positive

Components of Error

Error Components	Measurement in 1990	Measurement in 2000
P-sample matching error	1990 Matching Error Study	1990 Matching Error Study with adjustments for 2000
P-sample data collection error	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
P-sample fabrication	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
E-sample data collection error	1990 Evaluation Followup	1990 Evaluation Followup with adjustments for 2000
E-sample processing error	1990 Matching Error Study	1990 Matching Error Study with adjustments for 2000
Correlation bias	1990 Demographic Analysis	2000 Demographic Analysis
Ratio estimator bias	1990 PES	2000 A.C.E.
Sampling error	1990 PES	2000 A.C.E.
Imputation error	1990 Reasonable Alternatives Imputation Study	1990 Reasonable Alternatives with adjustments for 2000
Excluded Census Data Error	1990 Excluded Data Study	Not available
Contamination of P sample by enumeration or vice versa	Shown to be negligible	Not available in time for analysis for decision
Misclassification error of records into poststrata from inconsistent reporting	Not measured	Not available in time for analysis for decision
Synthetic error	Artificial population analysis and not integrated in total error model	Under development but will not be integrated in total error model

Assumptions and Limitations:

- The assumption for nonsampling error components from field and processing operations is that the errors measured in the 1990 PES scaled for the 2000 population reasonably reflect the errors for the 2000 A.C.E.
- The mapping of the 1990 PES poststrata to the 2000 A.C.E. poststrata uses characteristics from the 1990 census.



95% Confidence Intervals for UC Rate
(Some Imputation Error, Corr Bias = 0)
%
12
7
2
-3
-8
US
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Evaluation Poststrata 2-5-01

Confidence intervals for Evaluation Poststrata 2-5-01

EV	UC rate	confidence interval
US	1.1789	(0.1442 , 0.7753)
1	0.1878	(-0.6402 , 0.2444)
2	-2.8191	(-6.9092 , -3.6579)
3	1.284	(-0.2965 , 4.3229)
4	0.2127	(-0.7902 , 0.3845)
5	2.33	(0.229 , 3.5401)
6	0.42	(-1.6951 , -0.0173)
7	1.13	(0.0387 , 2.0865)
8	1.84	(-0.4912 , 2.9166)
9	2.59	(0.1564 , 2.8541)
10	1.33	(-0.5871 , 1.2406)
11	-0.678	(-3.699 , 0.1861)
12	0.77	(-1.9824 , 1.3399)
13	3.5	(1.9521 , 4.354)
14	4.2	(1.0679 , 4.6)
15	3.97	(0.0181 , 3.49)

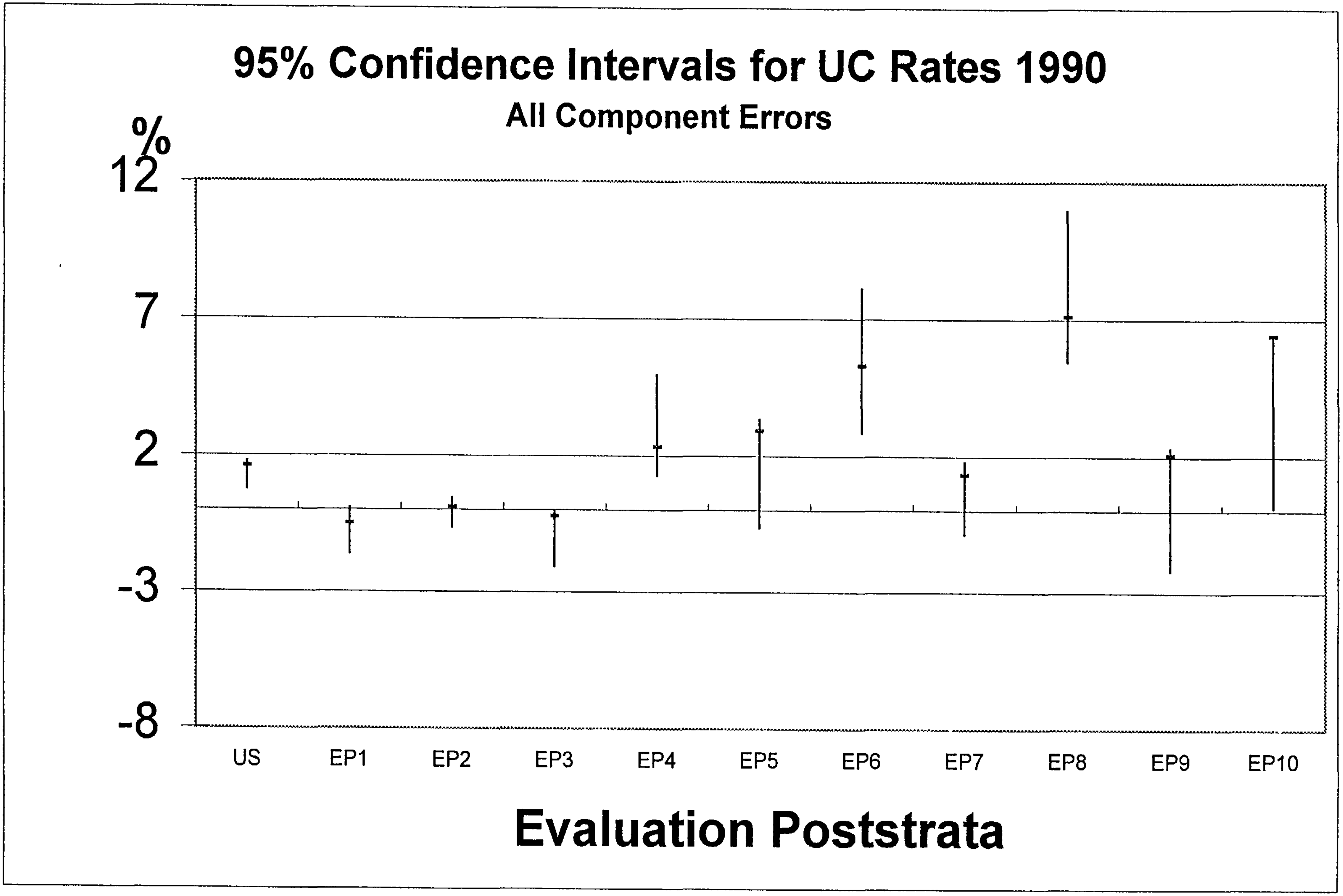
15 Evaluation Poststrata (2-2-01)

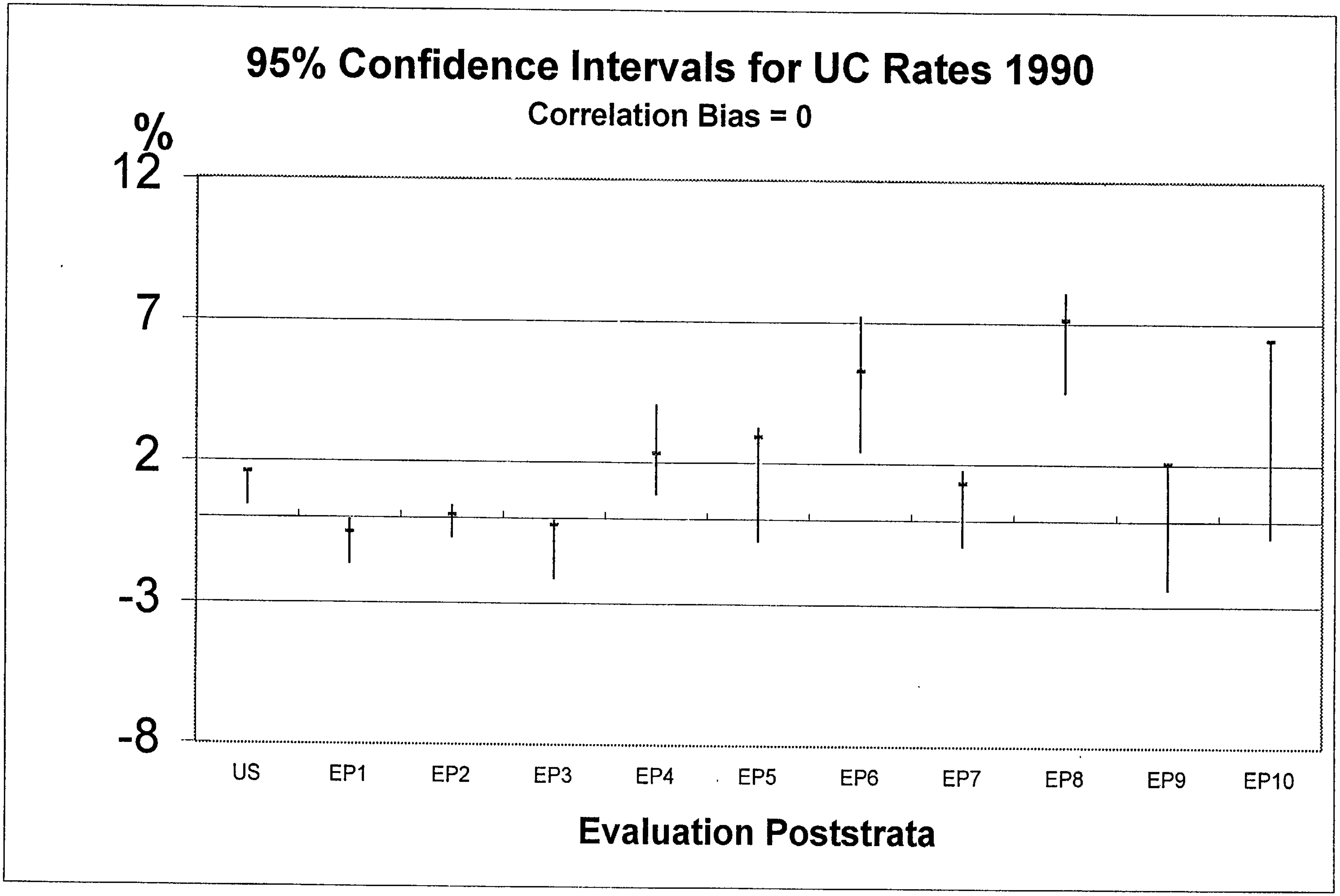
		No. in MVF P-sample (1990)	PS Groups (2000)
1. Non-minority/owner/large and Medium MSA MO-MB	high RR	12,662	1-4, 9-12
2. Non-minority/owner/large and Medium MSA MO-MB NE/MW	low RR	3,031	5,6,13,14
3. Non-minority/owner/large and Medium MSA MO-MB S/W	low RR	2,936	7,8,15,16
4. Non-minority/owner/Small MSA and Non-MSA MO-MB	high RR	5,560	17-20
5. Non-minority/owner/ Small MSA and Non-MSA MO-MB	low RR	2,095	21-24
6. Non-minority/Owner/All Other TEAs		7,355	25-32
7. Non-minority/ Non-Owner/Large or Medium MSA MO-MB	high RR	4,963	33, 35
8.Non-minority/ Non-Owner/Large or Medium MSA MO-MB	low RR	3,197	34, 36
9. Non-minority/non-owner/Small MSA & Non-MSA MO-MB All other TEA		5,291	37-40
10. Minority/owner/large and Medium MSA MO-MB	high RR	8,841	41, 49, 57,59
11. Minority/owner/large and Medium MSA MO-MB	low RR	5,628	42, 50
12. Minority/Owner/All Other TEAs		3,877	43, 44, 51, 52
13. Minority/ Non-Owner/Large or Medium MSA MO-MB	high RR	10,809	45, 53, 58,60
14. Minority/ Non-Owner/Large or Medium MSA MO-MB	low RR	6,421	46, 54
15. Minority/Non-Owner/All Other TEAs		3,797	47, 48, 55, 56, 61-64
Total		86,463	

Table 3: Census 2000 A.C.E. - 64 Post-Stratum Groups

Race/Hispanic Origin Domain Number*	Tenure	MSA/TEA	High Return Rate				Low Return Rate			
			N	M	S	W	N	M	S	W
Domain 7 (Non-Hispanic White or "Some other race")	Owner	Large MSA MO/MB	1	2	3	4	5	6	7	8
		Medium MSA MO/MB	9	10	11	12	13	14	15	16
		Small MSA & Non-MSA MO/MB	17	18	19	20	21	22	23	24
		All Other TEAs	25	26	27	28	29	30	31	32
	Non-Owner	Large MSA MO/MB	33				34			
		Medium MSA MO/MB	35				36			
		Small MSA & Non-MSA MO/MB	37				38			
		All Other TEAs	39				40			
Domain 4 (Non-Hispanic Black)	Owner	Large MSA MO/MB	41				42			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	43				44			
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	45				46			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	47				48			
		All Other TEAs								
Domain 3 (Hispanic)	Owner	Large MSA MO/MB	49				50			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	51				52			
		All Other TEAs								
	Non-Owner	Large MSA MO/MB	53				54			
		Medium MSA MO/MB								
		Small MSA & Non-MSA MO/MB	55				56			
		All Other TEAs								
Domain 5 (Native Hawaiian or Pacific Islander)	Owner		57							
	Non-Owner		58							
Domain 6 (Non-Hispanic Asian)	Owner		59							
	Non-Owner		60							
American Indian or Alaska Native — **Domain 1** (On Reservation)	Owner		61							
	Non-Owner		62							
American Indian or Alaska Native — **Domain 2** (Off Reservation)	Owner		63							
	Non-Owner		64							

* For Census 2000 persons can self-identify with more than one race group For post-stratification, persons are included in a single Race/Hispanic Origin domain. This does not change a person's actual response and all persons will be tabulated based on their actual response in the census.


95% Confidence Intervals for UC Rates 1990
All Component Errors
%
12
7
2
-3
-8
US
EP1
EP2
EP3
EP4
EP5
EP6
EP7
EP8
EP9
EP10
Evaluation Poststrata


95% Confidence Intervals for UC Rates 1990
Correlation Bias = 0
%
12
7
2
-3
-8
US
EP1
EP2
EP3
EP4
EP5
EP6
EP7
EP8
EP9
EP10
Evaluation Poststrata

ESCAP MEETING NO. 33 - 02/06/01

MINUTES

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 33
February 6, 2001

Prepared by: Nick Birnbaum

The thirty-third meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 6, 2001 at 10:30.

The agenda for the meeting was to continue discussions of the results from After Follow-up Person Matching and to examine preliminary results from the total error model analysis.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
Tommy Wright
Donna Kostanich
Raj Singh
David Whitford
William Bell
Danny Childers
Deborah Fenstermaker
Mary Mulry
Nick Birnbaum
Sarah Brady
Carolee Bush
Annette Quinlan
Kathleen Styles
Maria Urrutia

I. After Follow-up Person Matching Results

Danny Childers continued his presentation of the After Follow-up Person Matching results. Data on the type of erroneous enumeration as a percent of the total E-sample were compared with 1990 data. This comparison revealed, among other things, that the erroneous enumeration categories of: duplicate and "other residence" were a smaller percentage of the E-sample in 2000 than in 1990, while geocoding error was a larger percentage in 2000. The lower E-sample percentage of duplicates can be explained by the specific operations implemented in Census 2000 to identify and delete duplicated housing units from the Master Address File.

The Committee then examined data on the type of erroneous enumeration as a percent of the total E-sample resolved cases by different variables, including race domain, tenure, type of enumeration area (TEA), mail return rate, and region. It was noted that in mailout/mailback areas with low mail return rates, duplicates were a larger percent of the E-sample than for mailout/mailback areas with high return rates. This observation can be explained by the fact that those households that were identified as not responding to the mailout campaign were more likely to be duplicated in the address file, which would result in a lower mail return rate, thereby producing the observed relationship.

The Committee also observed that erroneous enumerations were a smaller percent of the E-sample in the Midwest region. Additionally, when examining the percent of E-sample for various types of erroneous enumerations by regional office, it was observed that the New York Regional Office area had among the highest percentages for insufficient information and duplicate cases. These observations led into a discussion about the synthetic assumption and how localized factors that affect coverage and bear no relationship to the post-stratification variables are not accounted for in the dual system estimation.

In summary, the P-sample and E-sample data presented were consistent with the Census Bureau's understanding of how well Census 2000 was conducted. There was some evidence of a regional effect for components of coverage error, which heightened concern about the synthetic assumption.

II. Overview of Preliminary Results for the Total Error Model

Mary Mulry presented preliminary results from the total error model for fifteen evaluation post-strata. These results do not include the effect of correlation bias and some imputation error. The attached information, based on the DSSD Census 2000 Procedures and Operations Memorandum Series Number B-13, presented a description of the evaluation post-strata, and confidence intervals for the A.C.E. estimates for these post-strata based on the bias and variance estimates from the preliminary total error analysis (excluding correlation bias). The Committee reviewed the data and noted evidence of a regional effect on the post-strata undercount rate estimates. The Committee further noted that additional examination and discussion were required to understand the full importance of these preliminary results. Additionally, it was noted that it was important to understand the relationship between the total error analysis and how it is used in the loss function analysis to compare the initial census and A.C.E. results.

III. Next Meeting

The agenda for the next meeting, to be held on February 7, 2001, is to compare the demographic analysis and A.C.E. estimates.

ESCAP MEETING NO. 34 - 02/07/01

AGENDA

Kathleen P Porter
02/07/2001 09:38 AM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc: J Gregory Robinson/POP/HQ/BOC@BOC, Alfredo Navarro/DSSD/HQ/BOC@BOC

Subject: ESCAP Agenda for 2/7

The agenda for the February 7 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. Preliminary DA/DSEs - Greg Robinson

2. Census Race Classifications - Greg Robinson

3. Preliminary Loss Function Results - Freddie Navarro

ESCAP MEETING NO. 34 - 02/07/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Presentation for ESCAP on Demographic Analysis (DA)
February 7, 2001

Comparison of Demographic Analysis and A.C.E. Results

SUMMARY OF FINDINGS

1) The DA and A.C.E. estimates both measure a reduction in net undercount in Census 2000 compared to 1990, but DA implies a much greater change.

> * both methods measure a dramatic reduction in the net undercount rates of children (ages 0-17), for Blacks and Nonblacks
>
> * both methods show a reduction in the net undercount rates of Black men and women (18+)

2) DA also finds a dramatic reduction in the net undercount rates of Nonblack men and women in Census 2000, compared to the rates of previous censuses. The A.C.E. estimates indicate no change or a slight increase in undercount rates for Nonblack adults as a group.

> * we are investigating to what extent the DA estimates are "too low" for Nonblacks due to understatement of immigration in the DA components of growth since 1990 (especially for Hispanics and Asians).

3) DA "expected"sex ratios for Black adults are much higher than both the census sex ratios and A.C.E. sex ratios, implying that A.C.E. is not capturing the high undercount rates of Black men relative to Black women (the well-known "correlation bias")

Presentation for ESCAP on Demographic Analysis (DA)
February 7, 2001

Comparison of Demographic Analysis and A.C.E. Results

Total Population(Table 1)

The Census 2000 count of 281.4 million is 1.8 million higher than the DA estimate of 279.6 million. Relative to DA, the difference implies a net overcount of 0.7 percent. This net coverage is dramatically different from that in the 1990 census, where the net undercount was 4.7 million or 1.8 percent.

The Census 2000 count is 3.3 million lower than the A.C.E. estimate of 284.7 million, giving a 1.2 percent net undercount. The A.C.E. net undercount is slightly lower than the PES estimate of 1.6 percent in 1990.

The DA and A.C.E. estimates both measure a reduction in net undercount in Census 2000 compared to 1990, but DA implies a much greater change.

Sex (Table 2)

DA estimates show a large reduction in net undercount of both males and females. The drop was greater for males, reducing the male-female differential.

The A.C.E. measured a relatively small reduction in net undercount of males and females; the male-female gap remained the same.

Sex and Age (Table 3)

The more detailed DA estimates for sex and age groups continue to reveal the pervasiveness of the change in coverage from 1990. The estimate of net undercount in 2000 is much lower for children (ages 0-17) and adult age groups of both sexes. The fall in net undercount is most notable for males and females aged 18-29 (we are assessing the extent to which this estimate is biased due to understatement of growth from 1990 to 2000 by DA).

The A.C.E. measured a reduction in net undercount of children. Unlike DA, however, the A.C.E. does not show an overall improvement in coverage for adult males and females.

Race and Sex (Table 4)

For DA estimates classified by race, two different sets are presented–1) Model 1 compares the 2000 DA estimates for Blacks with Census 2000 tabulations for persons who marked only the race circle for "Black", (2) Model 2 compares the 2000 DA estimates for Blacks with Census 2000 tabulations for persons who marked the race circle for "Black" and other race response circles (1 or more other circles).

Net undercount rates of Blacks are lower in 2000 than 1990 under both sets, with the reduction being much greater under the Model 2 set. The rates dropped for Black males and Black females. Net undercount rates are also appreciably lower in 2000 for Nonblack males and Nonblack females; in fact, net overcounts are indicated.

The A.C.E. also measures a large drop in the net undercount rate of Black males and Black females in Census 2000, compared to the undercount estimated by the PES in 1990. The A.C.E. results do not show an improvement in coverage for Nonblacks, unlike the DA findings.

Race, Sex, and Age (Table 5)

Compared to historical DA trends, the DA estimates for 2000 reveal a decline for all race-sex-age categories. The net coverage change is large for all Nonblack age-sex groups and for Blacks under Model 2.

The A.C.E. finds a large reduction in the net undercount rate of Black children and most Black adult age categories. For Nonblacks, the undercount rate in 2000 is lower than 1990 for children. For Nonblack adults, the 2000 rates are at or slightly above the 1990 rates.

Sex Ratios (Tables 6 and 7)

The DA "expected" sex ratios for Blacks are much higher than the corresponding census sex ratios for adult age groups, indicative of the higher undercount rate of Black men relative to Black women (Table 6). Note that these findings are the same whether using Model 1 or Model 2. The "gap" in the sex ratios for Nonblacks are much smaller, reflecting a much smaller male-female difference in estimated undercount rates.

The A.C.E. sex ratios for Blacks are similar to the census sex ratios and fall below the DA sex ratios. The "low" A.C.E. ratios for Blacks in 2000 imply that the A.C.E. understated the net undercount of adult Black men (the well-known "correlation bias"). As shown by the historical sex ratio comparisons in Table 7, indication of correlation bias (relative to DA) are consistently found in the results of previous coverage measurement surveys.

Next Steps

We continue to assess the uncertainty in the DA estimates, in particular, the magnitude of an understatement of 1990-2000 growth and its implications on specific age-sex-race categories. We are using "error of closure" estimates in this assessment.

We will present some illustrative subnational DA benchmarks of coverage, which inform the overall demographic findings.

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Table 1- Census Count, Demographic Analysis (DA) Estimate, and Accuracy and Coverage Evaluation (A.C.E.) Estimate for the U.S. Resident Population: 4-1-2000

	Count or Estimate
1. Census Count	281,421,906
2. D.A. Estimate	279,581,684
3. A.C.E. Estimate	284,683,785
Difference from Census:	
4. D.A. Estimate (=2-1)	(1,840,222)
5. A.C.E. Estimate (=3-1)	3,261,879
Percent Difference	
4. D.A. Estimate (=4/2*100)	-0.66
5. A.C.E. Estimate (=5/3*100)	1.15

Note: The DA estimates for ages under 65 are based on components of population change (births, deaths, legal immigration, and estimates of emigration and undocumented immigration). The DA estimates for ages 65 and over are based on 2000 Medicare data, adjusted for underenrollment.

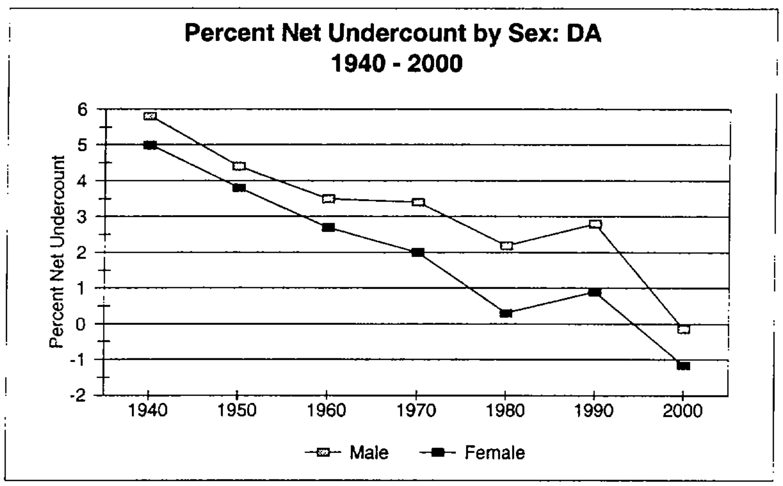
The A.C.E. and DA estimates are preliminary.

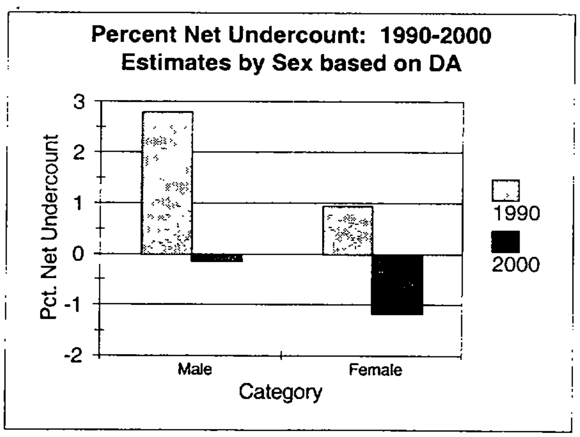
Table 2--Estimates of Percent Net Undercount by Sex: 1940 to 2000
(a minus sign denotes a net overcount)

	Demographic Analysis							Survey-based	
								PES	A.C.E.
Category	1940	1950	1960	1970	1980	1990	**2000**	1990	**2000**
Total Population	6.5	4.1	3.1	2.7	1.2	1.8	**-0.7**	1.6	**1.15**
Male	5.8	4.4	3.5	3.4	2.2	2.8	**-0.1**	1.9	**1.5**
Female	5.0	3.8	2.7	2.0	0.3	0.9	**-1.2**	1.3	**0.8**
Male:Female Diff.	0.8	0.6	0.8	1.4	1.9	1.9	**1.0**	0.7	**0.7**

Source: 1940-1990-- Robinson, J. Gregory, Bashir Ahmed, Prithwis Das Gupta, and Karen Woodrow, "Estimates of Population Coverage in the 1990 United States Census Based on Demographic Analysis", Journal of the American Statistical Association, Vol. 88, No. 423, pp. 1061-1077. Estimates for 2000 are unpublished preliminary results.

Source: 2000 - Preliminary A.C.E. and DA estimates. Universe is the U.S. resident population.





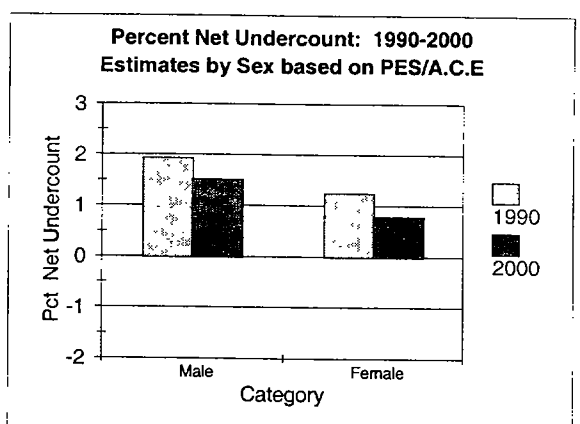
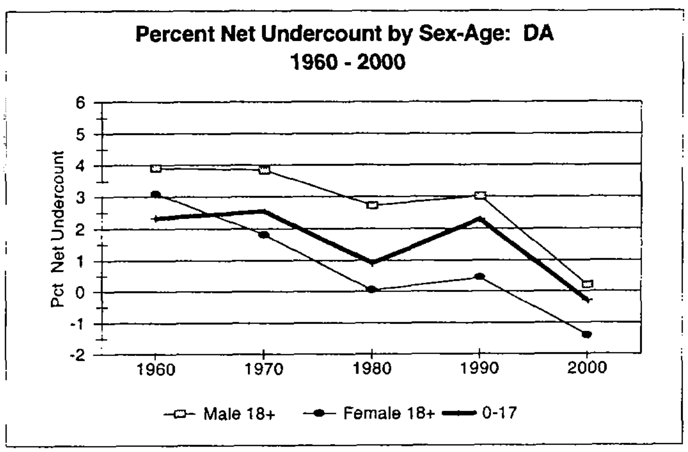
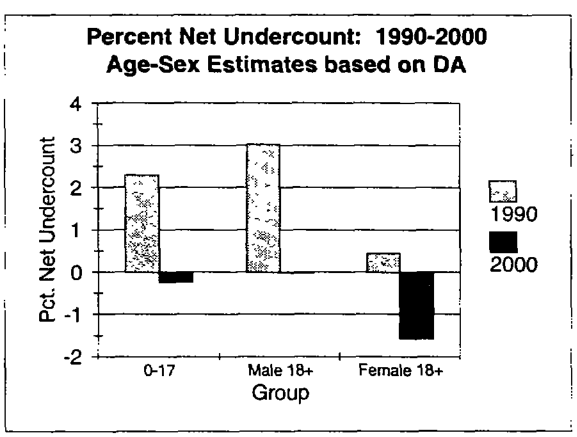


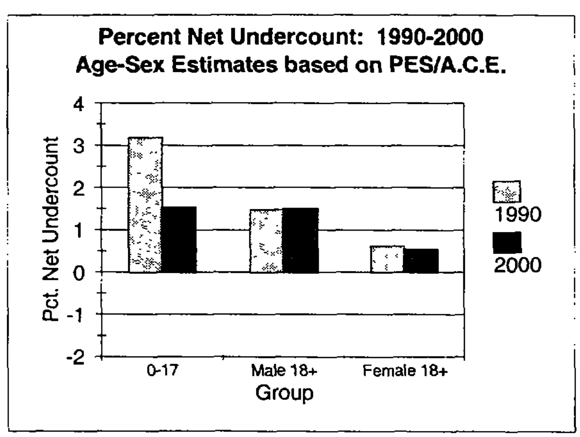
Table 3--Estimates of Percent Net Undercount by Sex and Age: 1960 to 2000
(a minus sign denotes a net overcount)

Category	Demographic Analysis					Survey-based	
						PES	**A.C.E.**
	1960	1970	1980	1990	**2000**	1990	**2000**
MALE							
Total	3.5	3.4	2.2	2.8	**-0.1**	1.9	**1.5**
0-17	2.8	2.7	0.9	2.2	**-0.5**	3.2	**1.5**
18-29	5.9	3.9	3.3	2.2	**-2.3**	3.2	**3.5**
30-49	4.2	5.1	3.6	3.8	**1.0**	1.9	**1.8**
50+	2.2	2.5	1.2	2.7	**0.3**	-0.6	**-0.2**
FEMALE							
Total	2.7	2.0	0.3	0.9	**-1.2**	1.3	**0.8**
0-17	1.8	2.4	0.9	2.4	**0.1**	3.2	**1.5**
18-29	2.8	1.3	0.4	0.6	**-3.1**	2.8	**2.1**
30-49	1.9	1.3	-0.0	0.5	**-0.9**	0.9	**1.0**
50+	4.6	2.6	-0.2	0.2	**-1.5**	-1.2	**-0.8**

Note: DA estimates are consistent with estimates in Table 2.







Revised 2-5-01

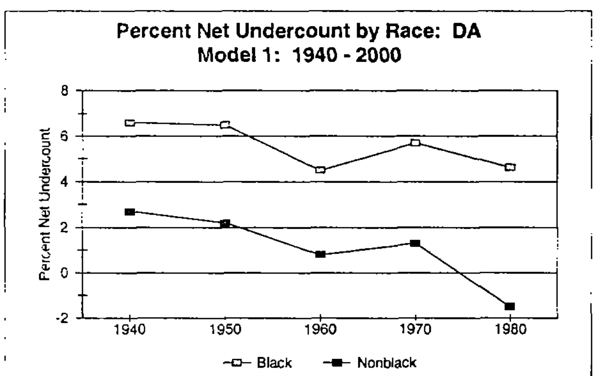
Table 4--Estimates of Percent Net Undercount by Race and Sex: 1940 to 2000
(a minus sign denotes a net overcount)

	Demographic Analysis						Survey-based	
					2000		PES	A.C.E.
Category	1960	1970	1980	1990	Model 1	Model 2	1990	2000
Total Population	3.1	2.7	1.2	1.8	**-0.7**	**-0.7**	1.6	**1.15**
Black	6.6	6.5	4.5	5.7	**4.6**	**0.9**	4.4	**2.1**
Male	8.8	9.1	7.5	8.5	**6.9**	**3.2**	4.9	**2.4**
Female	4.4	4.0	1.7	3.0	**2.5**	**-1.3**	4.0	**1.8**
Nonblack	2.7	2.2	0.8	1.3	**-1.5**	**-0.9**	1.2	**1.0**
Male	2.9	2.7	1.5	2.0	**-1.2**	**-0.7**	1.5	**1.4**
Female	2.4	1.7	0.1	0.6	**-1.7**	**-1.1**	0.9	**0.6**
Black:Nonblack Diff	3.9	4.3	3.7	4.4	**6.1**	**1.8**	3.3	**1.0**

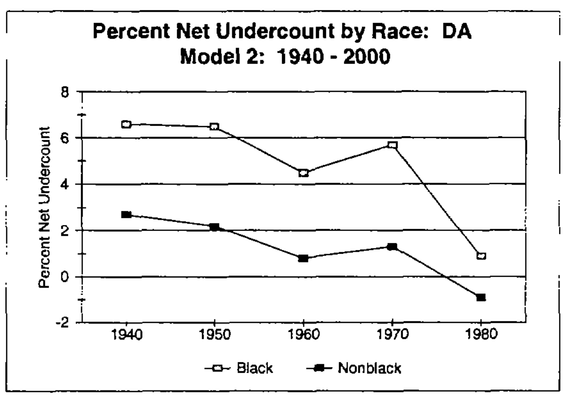
Note: Model 1 census tabulations for Blacks include persons who marked the Black circle and no other race response
Model 2 census tabulations for Blacks include persons who marked the Black circle and other response circles.
Persons who marked only the "Other race" circle are reassigned to a specific race category (to be consistent with 1990 DA estimates and the historical demographic data series)

Source: 1940-1990-- Robinson, J. Gregory, Bashir Ahmed, Prithwis Das Gupta, and Karen Woodrow, "Estimates of Population Coverage in the 1990 United States Census Based on Demographic Analysis", Journal of the American Statistical Association, Vol. 88, No. 423, pp. 1061-1077. Estimates for 2000 are unpublished preliminary results.

Source 2000 - See Table 2. Note that the A.C.E. estimates for Blacks pertain to the Non-Hispanic Blacks.

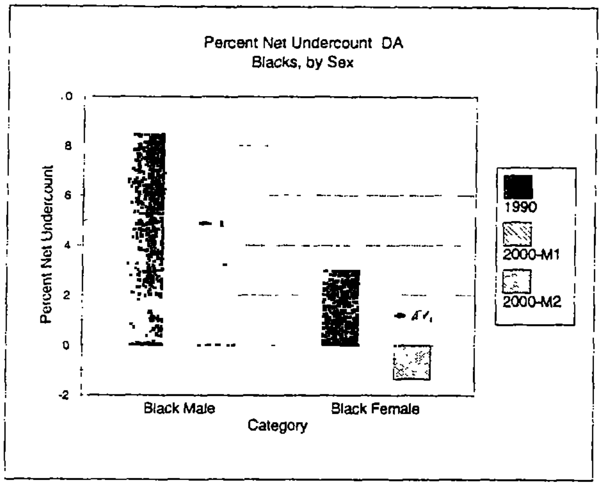


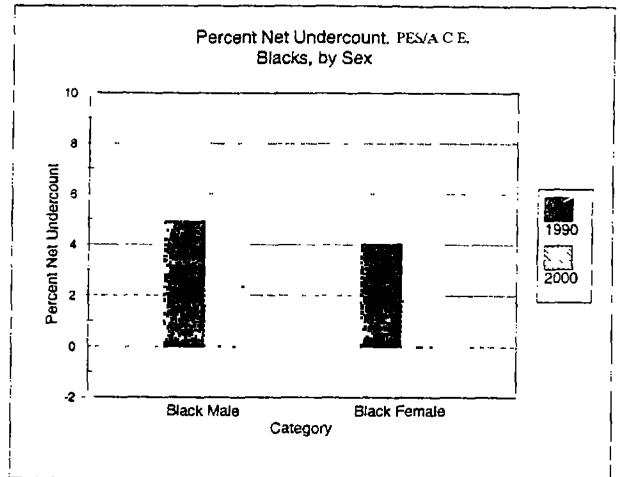
Percent Net Undercount by Race: DA Model 1: 1940 - 2000

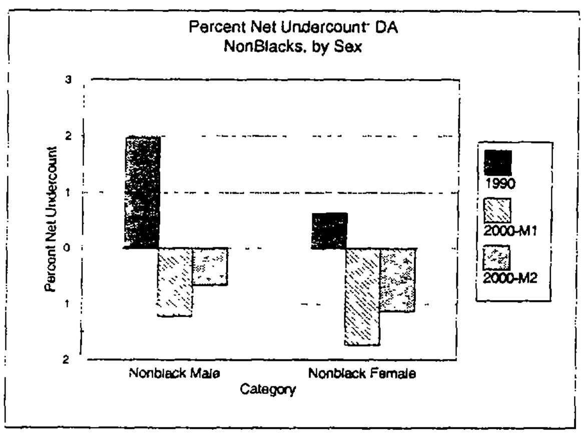


Percent Net Undercount by Race: DA Model 2: 1940 - 2000

Table 4. Continued







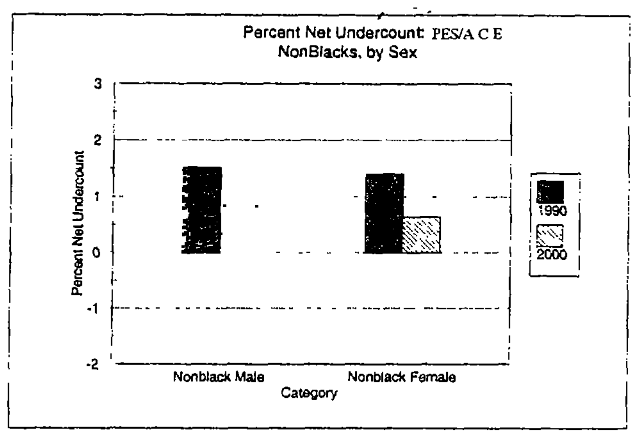
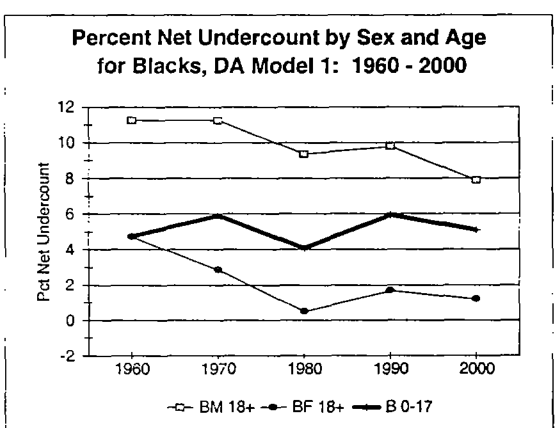


Table 5--Estimates of Percent Net Undercount by Race, Sex and Age: 1960 to 2000
(a minus sign denotes a net overcount)

	Demographic Analysis						Survey-based	
					2000		PES	**A.C.E.**
Category	1960	1970	1980	1990	Model 1	Model 2	1990	**2000**
BLACK MALE								
Total	8 8	9.1	7.5	8.5	**6.9**	**3.2**	4.9	**2.4**
0-17	5.4	6.2	4.2	5.9	**4.9**	**-1.9**	7.0	**3.0**
18-29	15.1	12.1	9.2	7.7	**8.1**	**4.9**	3.6	**3.7**
30-49	11.9	14.5	13.1	12.3	**10.2**	**8.3**	6.3	**2.6**
50+	6.6	6.3	4.6	8.3	**3.8**	**2.2**	-0.4	**-0.7**
BLACK FEMALE								
Total	4.4	4.0	1.7	3.0	**2.5**	**-1.3**	4.0	**1.8**
0-17	4.0	5.6	3.9	5.9	**5.4**	**-1.6**	7.1	**3.0**
18-29	5.4	4.5	2.4	2.9	**1.9**	**-1.7**	5.5	**3.8**
30-49	2.1	0.5	0.6	2.5	**2.1**	**-0.1**	3.2	**1.3**
50+	7.6	3.8	-1.9	-0.8	**-0.8**	**-2.5**	-1.2	**-0.8**
NONBLACK MALE								
Total	2.9	2.7	1.5	2.0	**-1.2**	**-0.7**	1.5	**1.4**
0-17	2.4	2.1	0.3	1.5	**-1.6**	**-0.2**	2.5	**1.3**
18-29	4.6	2.8	2.4	1.3	**-4.1**	**-3.5**	3.1	**3.4**
30-49	3.4	4.0	2.5	2.7	**-0.4**	**-0.1**	1.3	**1.7**
50+	1.8	2.2	0.9	2.2	**-0.1**	**0.1**	-0.6	**-0.2**
NONBLACK FEMALE								
Total	2.4	1.7	0.1	0.6	**-1.7**	**-1.1**	0.9	**0.6**
0-17	1.5	1.8	0.3	1.8	**-1.0**	**0.4**	2.5	**1.3**
18-29	2.4	0.9	0.1	0.3	**-4.0**	**-3.3**	2.4	**1.8**
30-49	1.9	1.3	-0.1	0.2	**-1.4**	**-1.0**	0.6	**0.9**
50+	4.3	2.5	-0.0	0.3	**-1.5**	**-1.3**	-1.2	**-0.8**

Sources and notes: See Table 2 and 4



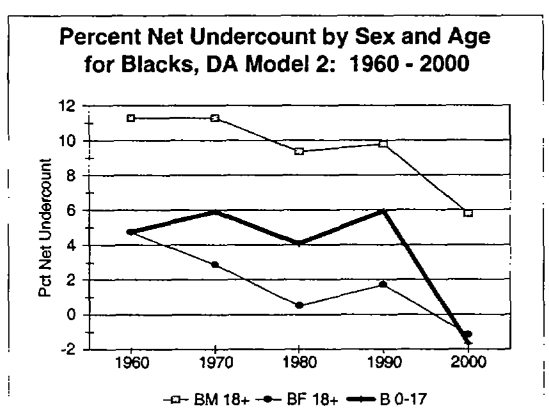
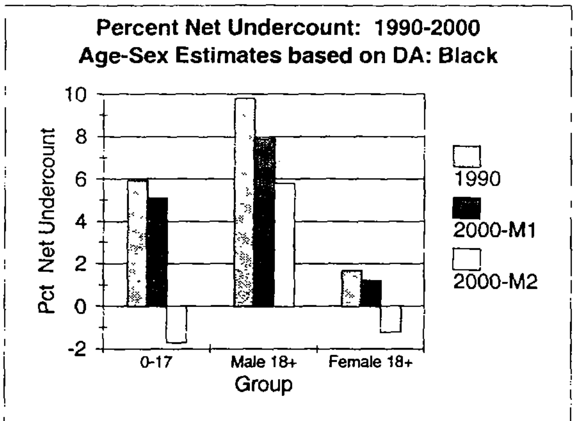
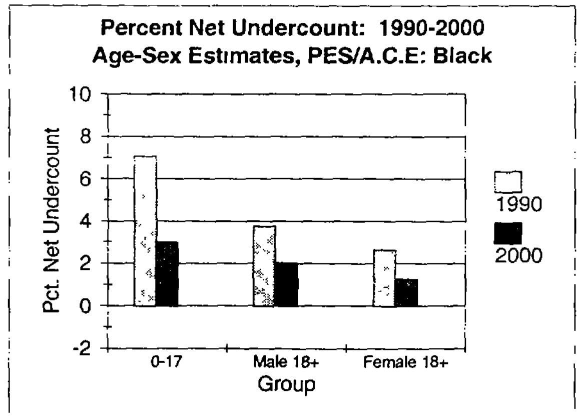
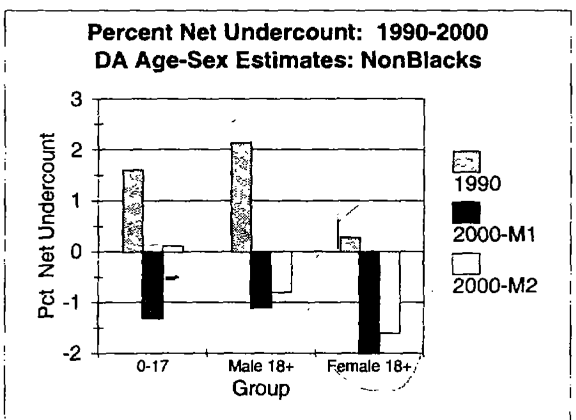
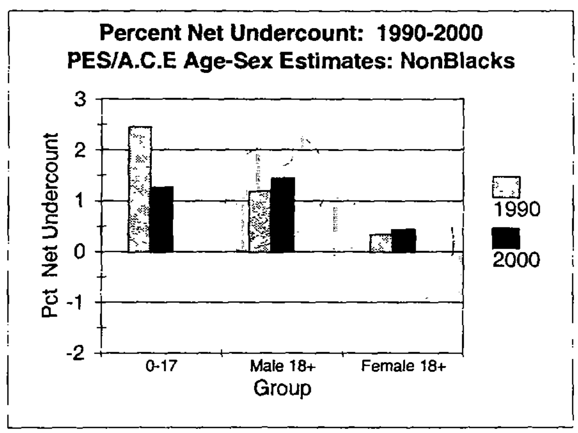


Table 5. Continued









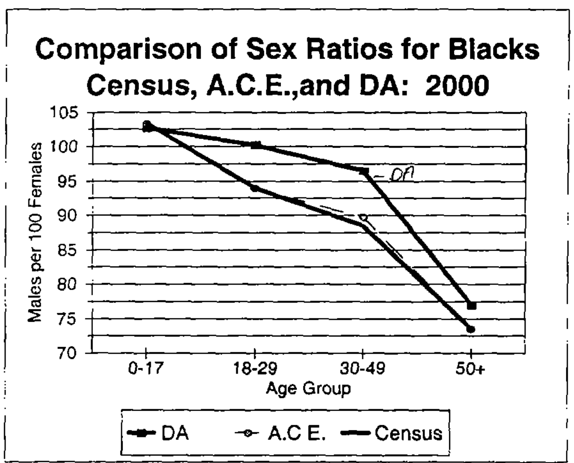
Revised 2-05-01

Table 6--Sex Ratios for the Census, Coverage Measurement Survey, and DA, by Race: 1990 and 2000

Category	1990 DA	1990 PES	1990 Census	2000 DA	2000 A.C.E. ~~Model 1~~	2000 Census Model 1	2000 Census Model 2
BLACK							
Total	95.2	90.4	89.6	94.9	91.1	90.6	90.7
0-17	102.4	102.4	102.4	102.7	103.3	103.3	103.1
18-29	99.3	92.1	94.0	100.2	93.9	94.0	93.7
30-49	95.9	89.0	86.2	96.5	89.7	88.5	88.4
50+	78.3	72.1	71.5	77.0	73.5	73.5	73.4
NONBLACK							
Total	97.2	96.5	95.9	97.6	97.9	97.1	97.2
0-17	105.2	105.5	105.5	105.0	105.5	105.5	105.6
18-29	104.9	104.6	103.8	105.2	106.9	105.3	105.4
30-49	102.0	100.3	99.6	101.6	101.4	100.6	100.6
50+	80.8	79.9	79.4	84.3	83.5	83.1	83.1

Note: Model 1 census tabulations for Blacks includes persons who marked the Black circle and no other response circle.
Model 2 census tabulations for Blacks includes persons who marked the Black circle and other response circles to the race question.

Source: DA, survey, and census data used to compute sex ratios are consistent with data used in Table 5.



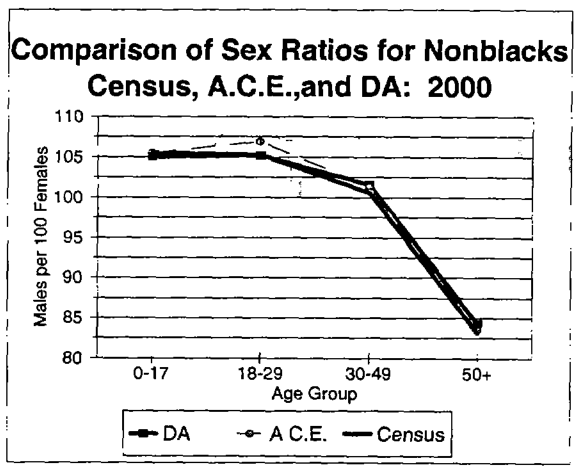
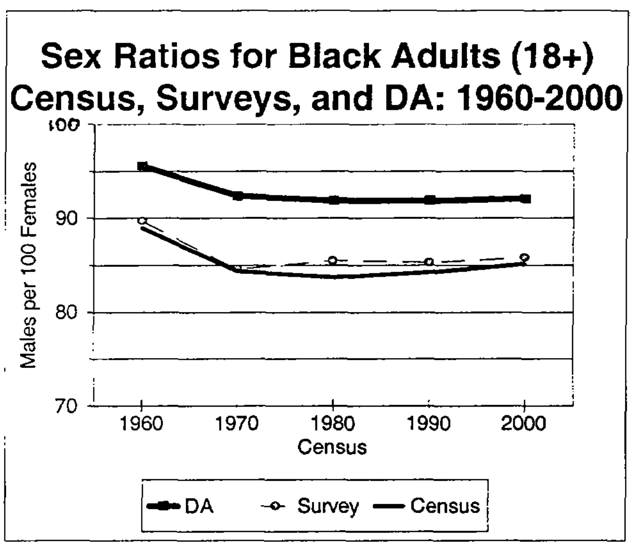


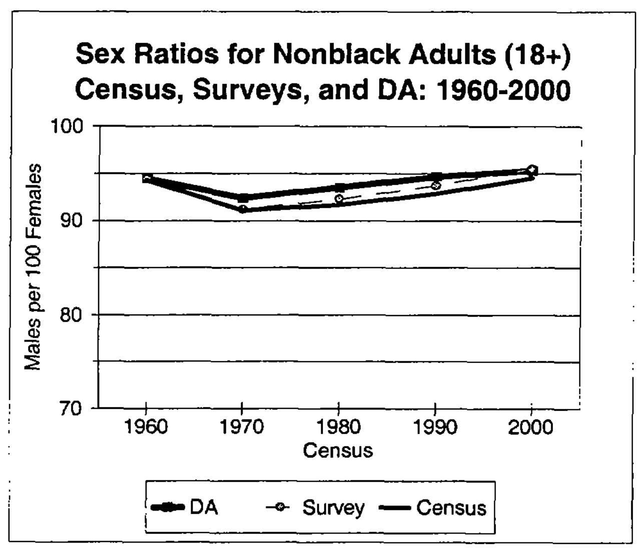
Table 7--Sex Ratios for the Census, Coverage Measurement Survey, and DA, for Adults by Race: 1960 - 2000

Category	1960	1970	1980	1990	2000	
					Model 1	Model 2
BLACK						
DA	95.6	92.4	91.9	91.9	92.1	92.1
Survey	89.7	84.6	85.5	85.3	85.8	85.8
Census	89.0	84.4	83.7	84.3	85.2	85.2
NONBLACK						
DA	94.5	92.4	93.5	94.7	95.3	95.3
Survey	94.4	91.2	92.3	93.7	95.5	95.5
Census	94.3	91.1	91.7	92.9	94.5	94.6

Note: Model 1 census tabulations for Blacks includes persons who marked the Black circle and no other circle. Model 2 census tabulations for Blacks includes persons who marked the Black circle and other response circles to the race question.

DA and census sex ratios refer to the population 18+ in all years. Survey estimates are 18+ in 1990 (PES). For 1980 (PEP), coverage rates by sex for the population 20+ were assumed to represent coverage of the population 18+; for 1970 and 1960 the available survey undercount estimates for 15+ are used.





ESCAP MEETING NO. 34 - 02/07/01

MINUTES

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 34

February 7, 2001

Prepared by: Annette Quinlan

The thirty-fourth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 7, 2001 at 10:30. The agenda for the meeting was to discuss the comparison of Demographic Analysis to the A.C.E. results.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines	Mary Mulry
Tommy Wright	David Word
Donna Kostanich	Roxie Jones
Raj Singh	Nick Birnbaum
William Bell	Kathleen Styles
Deborah Fenstermaker	Maria Urrutia
Alfredo Navarro	Sarah Brady
Gregg Robinson	Annette Quinlan

I. Introduction

Howard Hogan distributed a corrected table for the A.C.E. After Followup Matching results from the previous ESCAP meeting held on February 6, 2001. During that meeting a typing error was discovered in the table while the Committee was reviewing it.

II. Demographic Analysis

Gregg Robinson presented to the ESCAP a comparison summary between the A.C.E. results and the Demographic Analysis (DA) results. The handouts are attached. The main points, as summarized from the tables, are: (1) the DA and A.C.E. estimates both measure a reduction in net undercount in Census 2000 compared to 1990, (2) DA finds a dramatic reduction in the net undercount rates of Nonblack men and women in Census 2000 compared to the rates of previous censuses, and (3) the DA "expected" sex ratios for Black adults are much higher than both the census sex ratios and A.C.E. sex ratios.

Although the DA and A.C.E. both measure a reduction in the net undercount rate, there are important differences in coverage which have not yet been explained. The dramatic reductions in the DA net undercount rates of Nonblack men and women mentioned in (2) above are actually overcounts, while the A.C.E. estimates indicate no change from 1990 or a slight increase in the undercount rates for Nonblack adults as a group. These represent significant differences between the A.C.E. and DA. These differences need to be explained by the POP division and additional results will be presented at an ESCAP meeting next week.

The POP division staff are attempting to understand the causes of those differences in terms of how they measure immigration and emigration. There was also some discussion of uncertainties in the DA results due to the multi-race reporting for Census 2000. This is due to the fact that DA benchmarks are based on single race reporting while Census 2000 is based on multiple reporting causing variation between the two systems. This is even found in the 50+ age group, which is surprising because one would not expect a high level of variation within this age group.

It was requested that the field operations from Census 2000 be reviewed to determine if there is any explanation for duplicates or overcounts in the Census as a result of the data collection processes. John Thompson expressed concern that the apparent overcounts shown by the DA results were not detected in the A.C.E. results. The coverage measurement survey methodology, by design, has measured overcounts in the census very well, but why this trend seems to be reversed in the A.C.E. needs to be researched.

It was also noted that the effect on the A.C.E. estimates of whole person imputations and late adds in the census needs to be considered since they were a greater proportion of the census data than in 1990. DSSD staff will research this issue.

III. Next Meeting

The next meeting scheduled for Friday February 9, 2001 will discuss correlation bias and synthetic error methodology and results.

ESCAP MEETING NO. 35 - 02/09/01

AGENDA

Kathleen P Porter
02/09/2001 09:29 AM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc: William R Bell/SRD/HQ/BOC@BOC, Alfredo Navarro/DSSD/HQ/BOC@BOC

Subject: ESCAP Agenda for 2/9

The agenda for the February 9 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. Preliminary Loss Function Results - Freddie Navarro

2. Correlation Bias - Bill Bell

ESCAP MEETING NO. 35 - 02/09/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

	State Population	5% Error	Squared Error	Weighted Squared Error	Relative Squared Error
Wyoming	0.5	0.03	0.001	0.001	0.003
Rhode Island	1.0	0.05	0.003	0.003	0.003
Nevada	2.0	0.10	0.010	0.005	0.003
South Carolina	4.0	0.20	0.040	0.010	0.003
North Carolina	8.0	0.40	0.160	0.020	0.003
Florida	16.0	0.80	0.640	0.040	0.003
California	34.0	1.70	2.890	0.085	0.003
Total	281,421,906				

	State Share	5% Error (*1000)	Squared Error	Weighted Squared Error	Relative Squared Error
Wyoming	0.18%	0.09	0.008	4.4	2500
Rhode Island	0.37%	0.19	0.035	9.3	2500
Nevada	0.71%	0.36	0.126	17.8	2500
South Carolina	1.43%	0.71	0.508	35.6	2500
North Carolina	2.86%	1.43	2.045	71.5	2500
Florida	5.68%	2.84	8.063	142.0	2500
California	12.04%	6.02	36.216	300.9	2500
Total	100.00%				

	Estimated Size		True Size	Error	Sq Error	Sum of Errors	Weighted Sum of errors
State 1							
	504039	0.100649	0.10	0.001	0.000000		
	510054	0.10185	0.10	0.002	0.000003		
	498628	0.099569	0.10	0.000	0.000000		
	501786	0.100199	0.10	0.000	0.000000		
	489005	0.097647	0.10	-0.002	0.000006		
	494917	0.098828	0.10	-0.001	0.000001		
	507006	0.101242	0.10	0.001	0.000002		
	501965	0.100235	0.10	0.000	0.000000		
	488223	0.097491	0.10	-0.003	0.000006		
	512252	0.102289	0.10	0.002	0.000005		
						0.000015	
	5007875						363,820,439
State 2							
	501802	0.199295	0.20	-0.0007	0.000000		
	492861	0.195744	0.20	-0.0043	0.000018		
	503768	0.200076	0.20	0.0001	0.000000		
	508252	0.201857	0.20	0.0019	0.000003		
	511203	0.203029	0.20	0.0030	0.000009		
						0.00	
	2517886						198,031,263
State 3							
	511300	0.510677	0.50	0.0107	0.000114		
	489920	0.489323	0.50	-0.0107	0.000114		
						0.00	
	1001220						228,552,200

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Figure 1. 95% Confidence Intervals for 2000 A.C.E. Undercount Rates
Major Demographic Groups

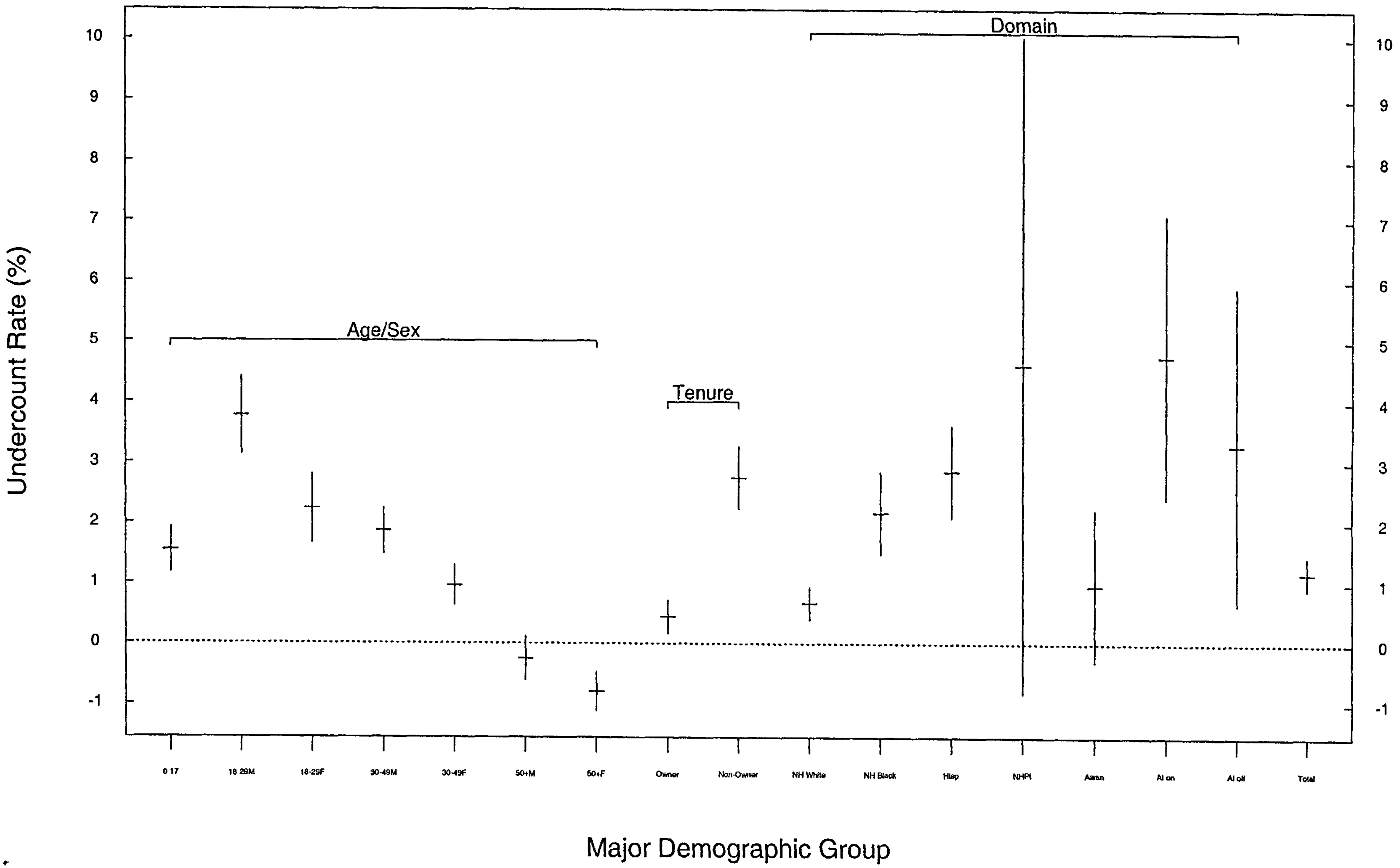


Figure 2. 95% Confidence Intervals for 2000 A.C.E. Undercount Rates Domain by Tenure

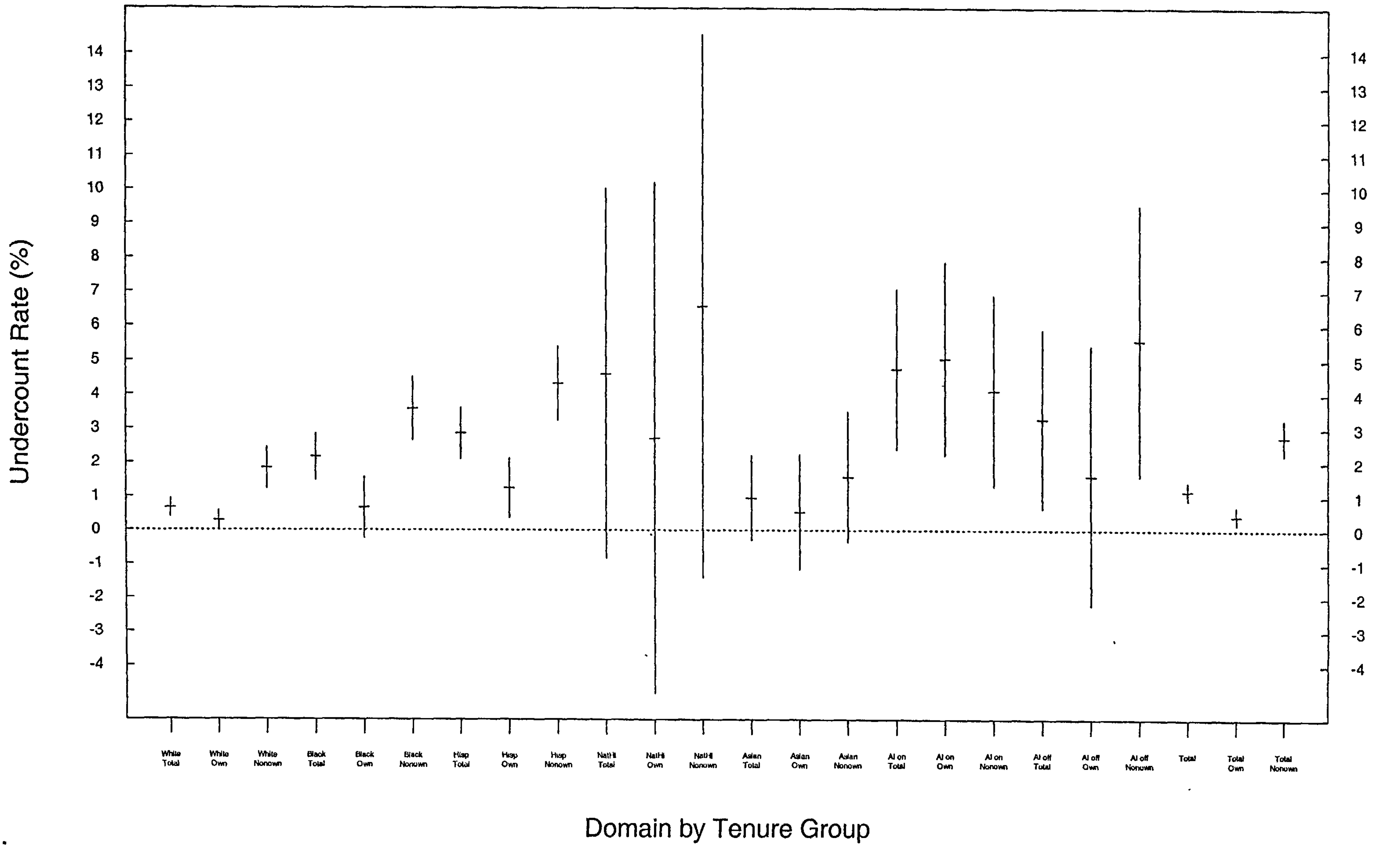


Figure 3. 95% Confidence Intervals for 2000 A.C.E. Undercount Rates
Tenure by Age/Sex

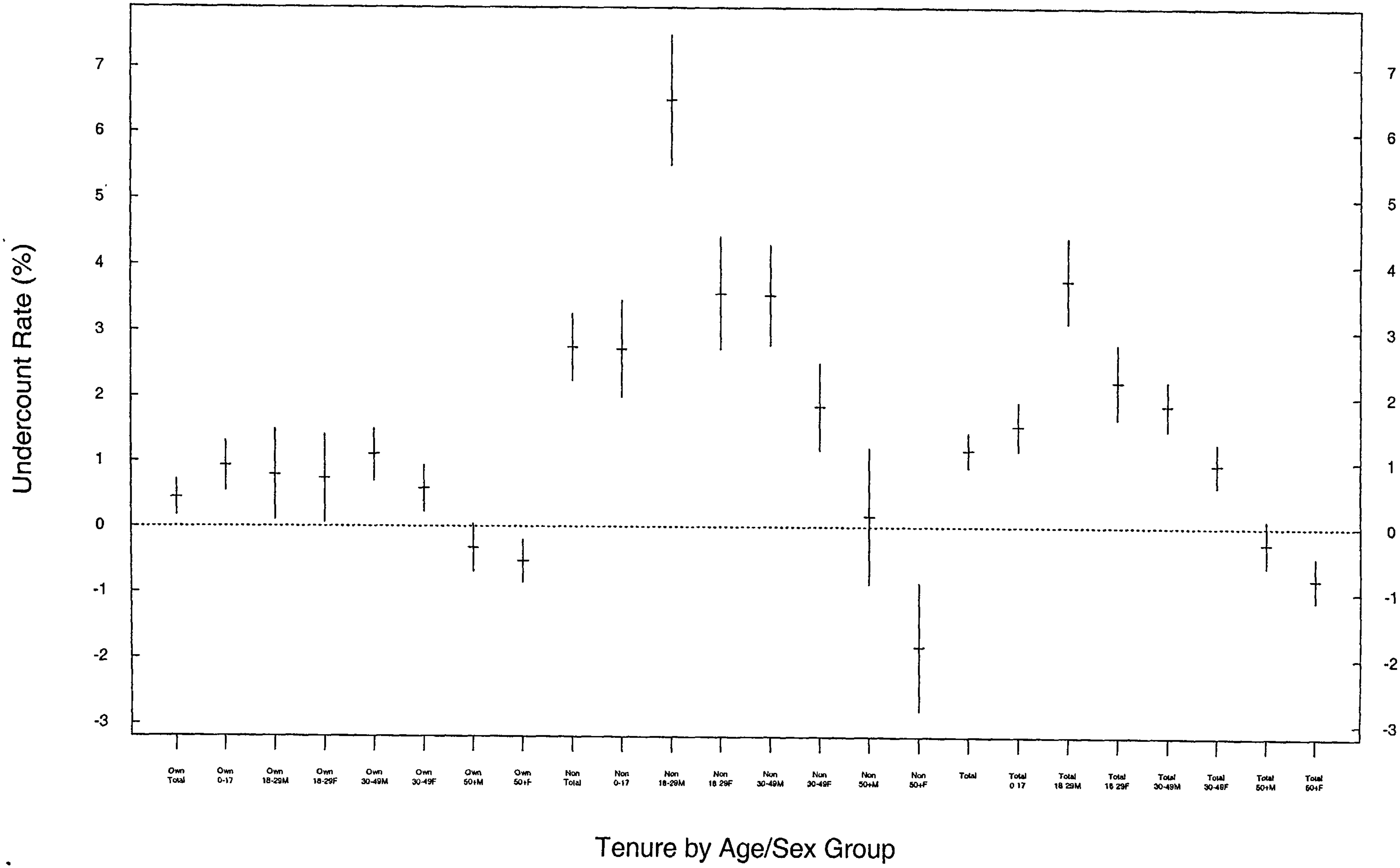


Figure 4. 95% Confidence Intervals for 2000 A.C.E. Undercount Rates Domain by Age/Sex

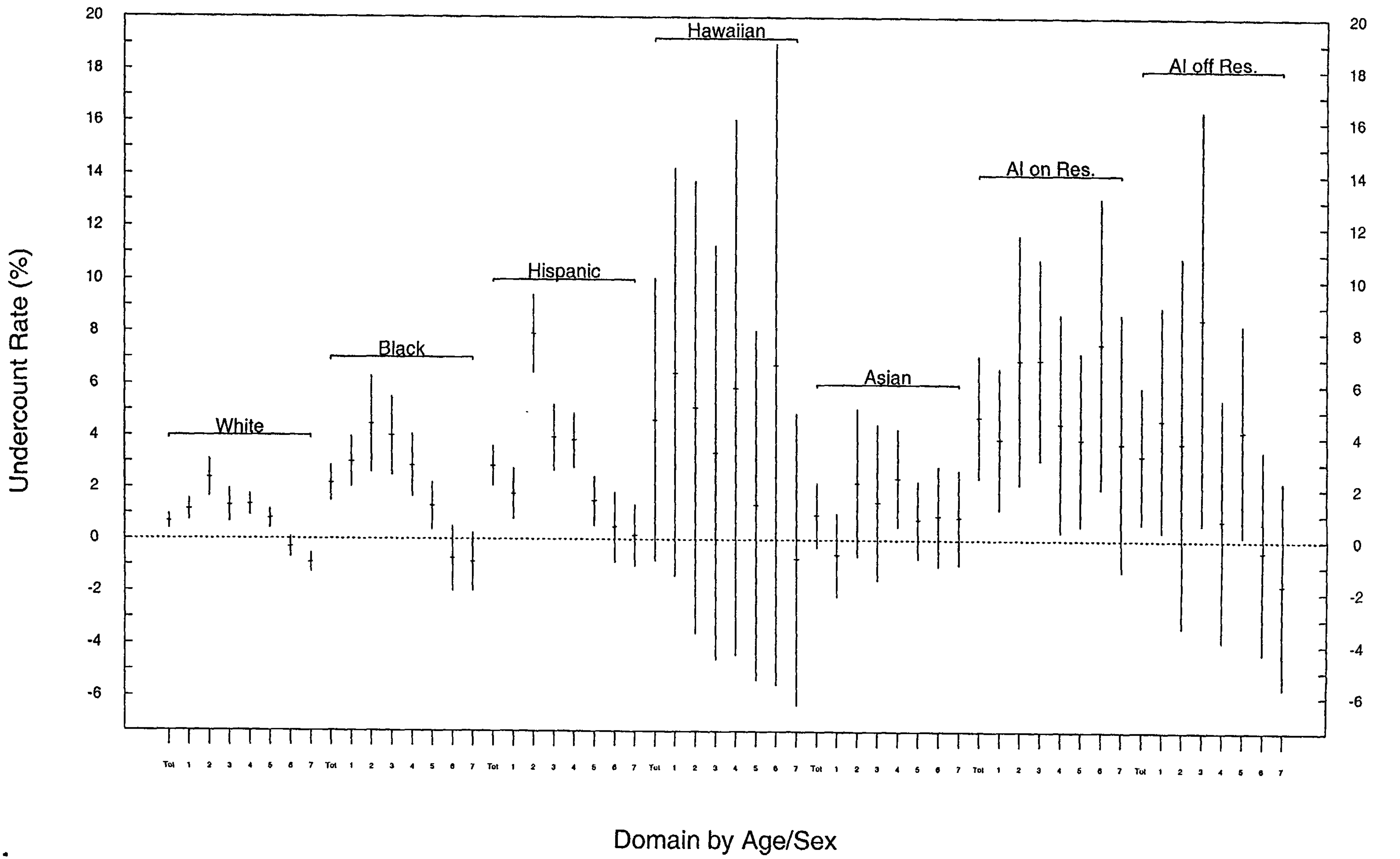
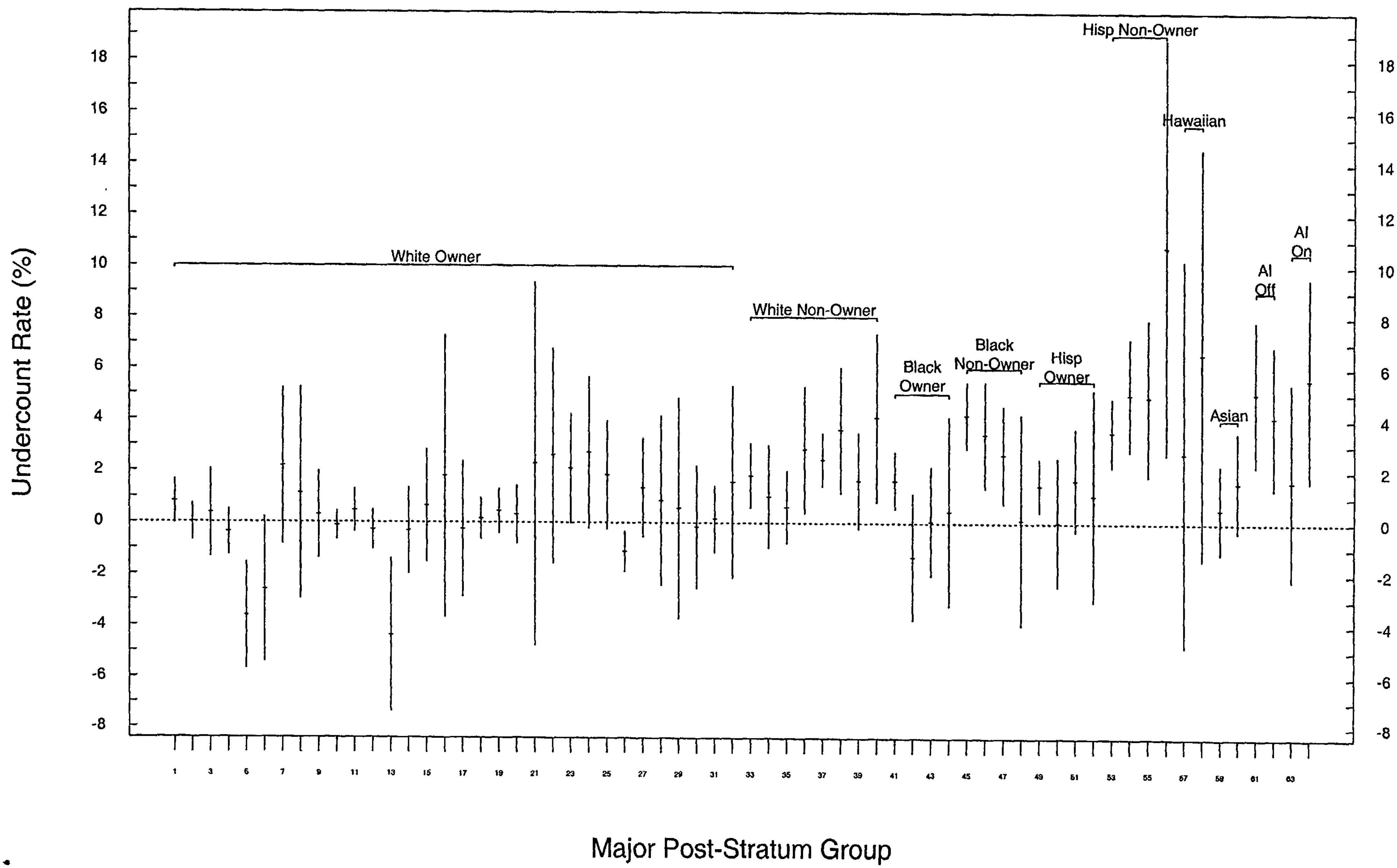


Figure 5. 95% Confidence Intervals for 2000 A.C.E. Undercount Rates
64 Major Post-Stratum Groups



Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Draft - February 09, 2001

Preliminary Loss Functions Results

- **Purpose of Loss Functions**

  To evaluate the accuracy of population counts or shares for the census and the corrected census.

- **Unit of Analysis**

  1. Congressional Districts
  2. States (including the District of Columbia)

- **Is the A.C.E. variance too large to prevent an adjustment of the census?**

  Assumption - We assume the A.C.E. is unbiased. The only error considered in this analysis is sampling error associated with the A.C.E..

- **Loss Functions**

  For the purpose of this analysis we use loss functions of the following form (used in 1990 loss function analysis):

$$CensusLoss = \sum_{i=1}^{n} W_i\left(Cen_i - T_i\right)^2$$

$$ACELoss = \sum_{i=1}^{n} W_i\left(ACE_i - T_i\right)^2$$

  W_i is a weight assigned to area i.
  T_i is the target or an estimate of the true population for area i.

  CensusLoss is a function of the census error and provides an aggregate measure of census inaccuracy across areas, for example: states and congressional districts.

  ACELoss is a function of the ACE error and similarly provides an aggregate measure of the ACE inaccuracy across areas.

- **What is compared?**

  The difference between the loss incurred by using the census and loss from using the corrected census figures.

CensusLoss - ACELoss = D.

If D<0, then the implication is that the census is more accurate than the ACE and adjusting the census may not be feasible. On the other hand, if D>0, then the ACE is more accurate.

Loss Functions

Type of Loss Functions	Census Loss	A.C.E. Loss
1. Squared Error Loss	$\sum_i \left(Cen_i - T_i\right)^2$	$\sum_i \left(ACE_i - T_i\right)^2$
2. Weighted Squared Error Loss	$\sum_i \left(Cen_i - T_i\right)^2 / Cen_i$	$\sum_i \left(ACE_i - T_i\right)^2 / ACE_i$
3. Relative Squared Error Loss	$\sum_i \left(Cen_i - T_i\right)^2 / Cen_i^2$	$\sum_i \left(ACE_i - T_i\right)^2 / ACE_i^2$
4. Equal CD Squared Error Loss (Only for Districts)	$\sum_j Cen_j^2 \sum_i \left(Cen_i - T_i\right)^2$	$\sum_j Cen_j^2 \sum_i \left(ACE_i - T_i\right)^2$

Table 1. U.S. Summary - Loss Functions Results

Type of Accuracy/Loss Function by Level	Census Loss (1)	ACE Loss (2)	CenLoss - ACELoss (3)
Distributive:			
Squared Error			
States	0.43931	0.14467	0.29464
Congressional Districts	29.04524	28.82922	0.21601
Weighted Squared Error			
States	**9.43157**	**4.85939**	**4.57217**
Congressional Districts	270.22660	185.00606	85.22054
Relative Squared Error			
States	529.10084	514.39830	14.70253
Congressional Districts	4,678.55000	2,353.81000	2,324.74000
Equal CD Squared Error			
States	NA	NA	NA
Congressional Districts	**1,864,608,874**	**972,680,795**	**891,928,078**

* Error reflects only A.C.E. variance and assumes an unbiased DSE.

+ The values for states and congressional districts corresponding to loss functions 1,2, and 3 are multiplied by 10^6.

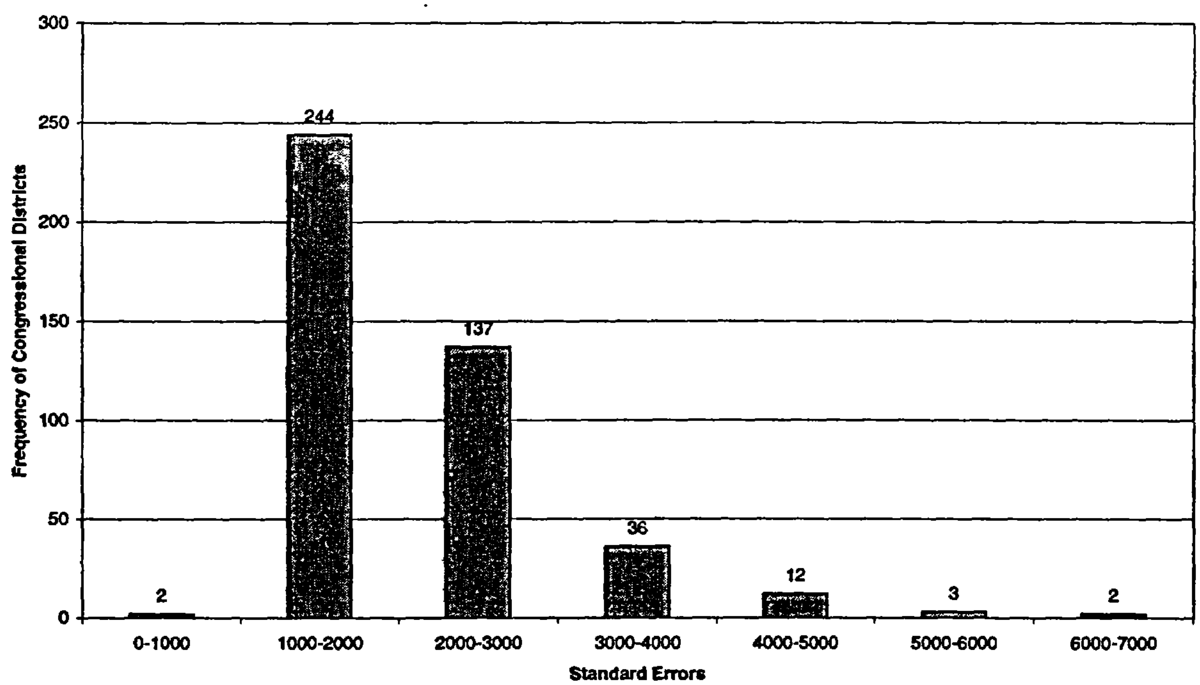
Table 2. U.S. Summary - Loss Functions Results

Type of Accuracy/Loss Function by Level	Census Loss (1)	ACE Loss (2)	CenLoss - ACELoss (3)
<u>Numeric</u>:			
Squared Error			
States	581,265,514,817	18,578,834,516	562,686,680,301
Congressional Districts	28,432,412,101	2,275,277,155	26,157,134,946
Weighted Squared Error			
States	**39,935**	**1,897**	**38,038**
Congressional Districts	42,943	3,418	39,525
Relative Squared Error			
States	0.0076	0.0006	0.0070
Congressional Districts	0.0658	0.0052	0.0606

* Error reflects only A.C.E. variance and assumes an unbiased DSE.

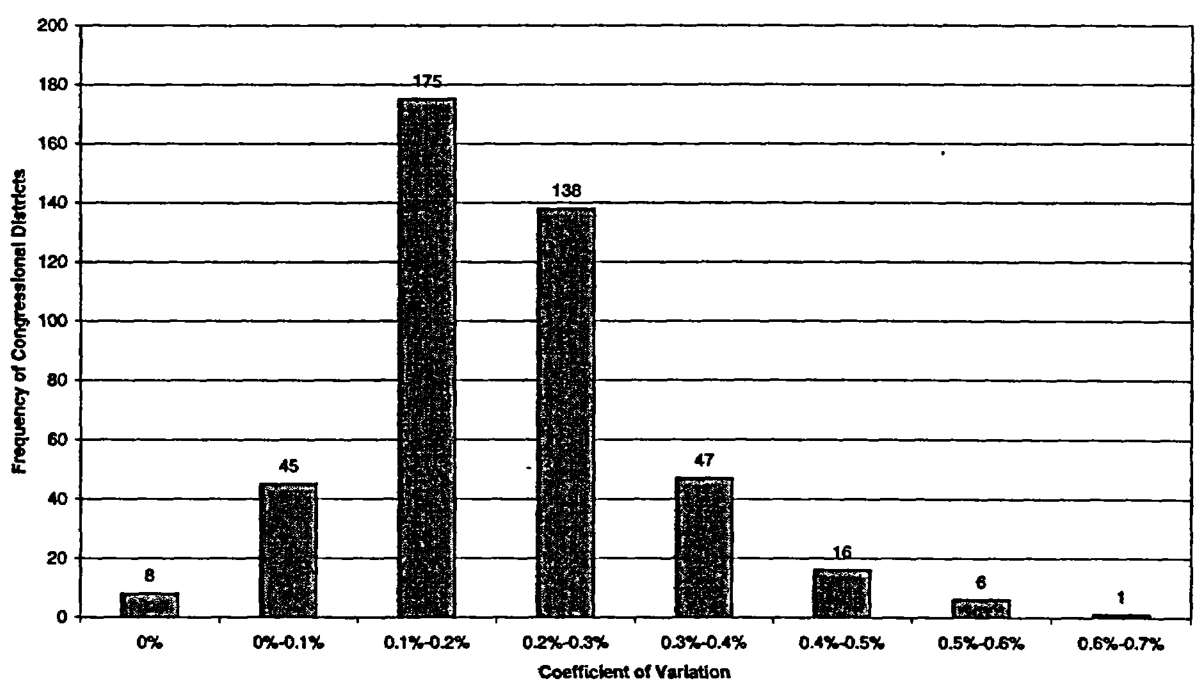
Graph I

Distribution of Standard Errors of A.C.E. Estimates of Congressional District Population



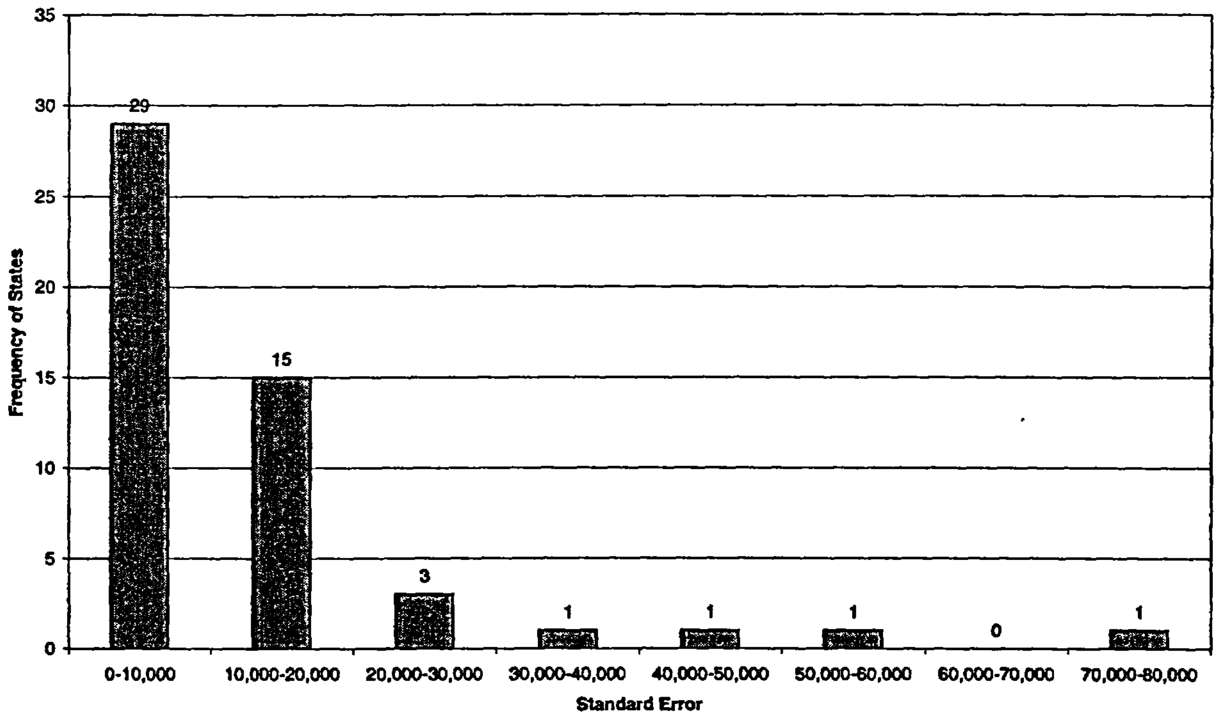
Graph II

Distribution of Coefficients of Variation of A.C.E. Estimates of Congressional District Shares



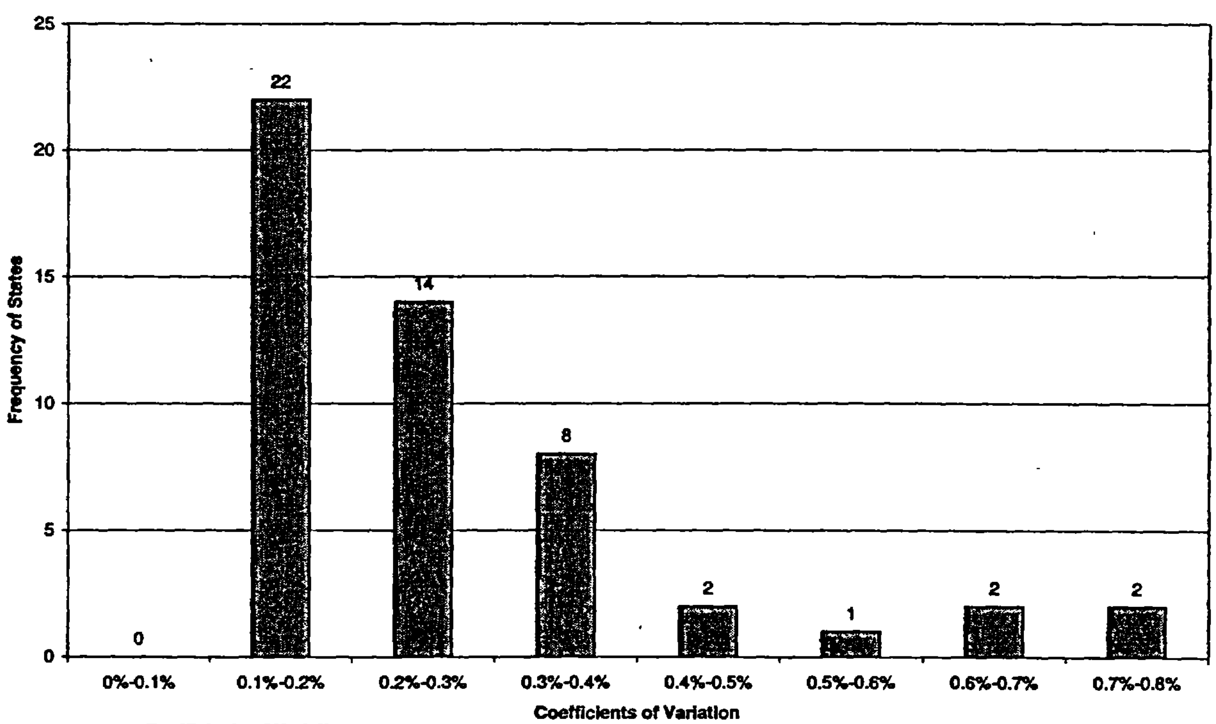
Graph III

Distribution of Standard Errors of A.C.E. Estimates of State Population

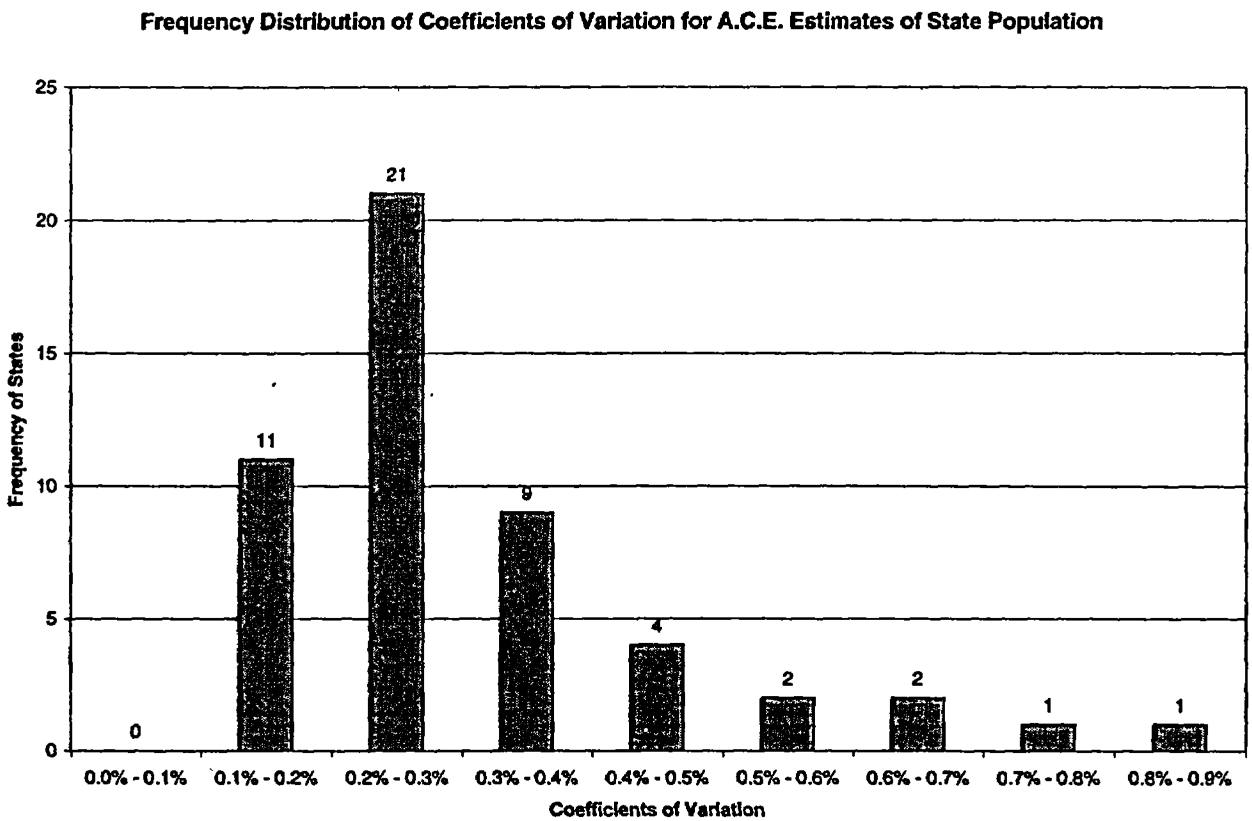


Graph IV

Distribution of Coefficients of Variation of A.C.E. Estimates of State Shares



Graph V


Frequency Distribution of Coefficients of Variation for A.C.E. Estimates of State Population
25
20
15
10
5
0
Frequency of States
0
11
21
9
4
2
2
1
1
0.0% - 0.1%
0.1% - 0.2%
0.2% - 0.3%
0.3% - 0.4%
0.4% - 0.5%
0.5% - 0.6%
0.6% - 0.7%
0.7% - 0.8%
0.8% - 0.9%
Coefficients of Variation

ESCAP MEETING NO. 35 - 02/09/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 35

February 9, 2001

Prepared by: Maria Urrutia and Sarah Brady

The thirty-fifth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on Friday, February 9, 2001 at 10:30. The agenda for the meeting was to present preliminary loss function results.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Ruth Ann Killion
Howard Hogan
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
Tommy Wright
Donna Kostanich
Raj Singh
William Bell
Alfredo Navarro
Deborah Fenstermaker
Nick Birnbaum
Kathleen Styles
Sarah Brady
Annette Quinlan
Maria Urrutia

I. Preliminary Loss Function Results

Howard Hogan presented examples of Loss Functions to illustrate how they are calculated and what they are intended to measure. Howard described two types of loss functions, weighted squared error and relative squared error, in his example, which is attached. A weighted square error loss function treats any potential loss or gain the same for an individual regardless of the area in which the individual lives. A relative squared error model treats entities, such as states or counties, equally. There are two types of accuracy that will be considered, numeric and distributive. Numeric accuracy compares entities' populations, while the distributive accuracy examines entities' shares of the total population. The Committee discussed the importance of analyzing these various measures of accuracy.

Alfredo Navarro then presented the preliminary loss functions results to the Committee. The preliminary results presented only included sampling error. The purpose of the presentation was a preliminary check to determine whether the change based on the A.C.E. was larger than sampling error. If the change between the A.C.E. and the census was less than the sampling error, the Committee would be able to conclude that the A.C.E. results should not be used.

Alfredo presented four loss functions for both types of accuracy for states and congressional districts. The loss function that was developed for congressional districts is a weighted measure, similar to the weighted squared error. The Loss Function formulas are attached. The Committee asked for detailed documentation of the loss functions and the assumptions. These documents will be provided to the Committee for their further analysis and review. All four of the loss functions presented indicated that the change between the A.C.E. and the census was greater than sampling error. Bob Fay noted that had the sample size for the A.C.E. been similar to 1990, the effect of the change would have been obscured by the sampling error.

Alfredo also presented the coefficients of variance resulting from the simulation to support the loss functions. The Committee asked for the coefficients of variance for the A.C.E. to be presented with the coefficients of variance for the simulation. These data are needed to confirm that the simulation coefficients of variance are in line with the A.C.E. estimates.

II. Graphical Representation of 95 percent Confidence Intervals for Undercount Rates

Donna Kostanich distributed graphics of the 95 percent confidence intervals for the undercount rates based on the Dual System Estimates (DSEs). Graphs of the 90 percent confidence intervals for the undercount rates had been presented at the February 2, 2001 ESCAP meeting and the Committee requested 95 percent confidence intervals be used for all their analyzes even though the Bureau's standard is 90 percent confidence intervals.

III. Next Meeting

The agenda for the next meeting, scheduled for Monday, February 12, 2001, is correlation bias.

ESCAP MEETING NO. 36 - 02/12/01

AGENDA

Kathleen P Porter
02/09/2001 03:37 PM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carol M Van Horn/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sarah E Brady/DMD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Vanessa M Leuthold/DMD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC

cc: Jacqueline M Cusick/DIR/HQ/BOC@BOC, William R Bell/SRD/HQ/BOC@BOC

Subject: ADDITIONAL ESCAP MEETING SCHEDULED FOR MONDAY!!

There will be an additional ESCAP meeting on Monday Feb. 12 from 2:00-3:30 p.m. in Rm. 2412/3.

The agenda topic will be:

Correlation Bias - Bill Bell

ESCAP MEETING NO. 36 - 02/12/01

HANDOUTS

Materials attached to these minutes were draft and preliminary material to inform the ESCAP Committee. The data and analysis contained in these documents are subject to revision and are not final. These materials report the results of research and analysis undertaken by Census Bureau staff. They have undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau.

Estimating Correlation Bias in A.C.E. 2000 Estimates

William R. Bell, room 3230-4, Bureau of the Census
ph: 301-457-4728

Revised February 13, 2001

What is "correlation bias"?

- Correlation bias can be defined as the error in poststratum estimates of population from DSEs based on "perfect data" (no sampling error, no other biases).
- Correlation bias can be expressed in alternative ways. If N is the true poststratum population and $\widetilde{\text{DSE}}$ the DSE based on perfect data, we could write:
  1. arithmetic correlation bias $= E(\widetilde{\text{DSE}}) - N$
  2. relative correlation bias $= [E(\widetilde{\text{DSE}}) - N] / N$
  3. correlation bias relative to census undercount
  4. correlation bias relative to expected $(2, 2)$ cell
- In 1990 we used definition 4. This time we are using 2.
- The existence of correlation bias is not, by itself, a valid criticism of adjustment. Since we would generally expect correlation bias, when present, to lead towards underestimation, in the absence of other errors the following relation would generally hold
$$\text{Census} < E(\text{DSE}) < \text{ truth.}$$

What produces correlation bias?

Failure of the general independence assumption underlying DSE due to

- **causal dependence**—the act of being included in the census makes someone more likely or less likely to be included in A.C.E., and/or
- **heterogeneity**—census and A.C.E. inclusion probabilities vary over persons within poststrata.

If heterogeneity exists in the sense that those more likely to be included in the census are also more likely to be included in A.C.E., then *correlation bias is negative (DSE underestimates)*

What evidence do we have of correlation bias in DSEs?

Comparisons against estimates from demographic analysis (DA) for age-sex-race groups (Black vs. Nonblack race) at the national level.

- Comparisons against DA totals—see Table 1.
- Comparisons against DA sex ratios—see Table 2.

Conclusions from comparisons against DA:

- large discrepancy between 2000 A.C.E. and DA results for Nonblacks 18-29, for both males and females
- evidence of significant correlation bias for adult Black males
- little evidence of correlation bias for females or children
- possibility of correlation bias for adult Nonblack males

How can we estimate correlation bias in DSEs (for adult males)?

DA totals are not sufficiently reliable $\Rightarrow$ use DA sex ratios.

1. Assume no correlation bias for adult females.
2. Aggregate DSEs for females to national level within age-race groups (Black vs. Nonblack). Multiply the female totals by DA sex ratios to get control totals for males.
3. Assume some measure of correlation bias (parameter) is constant over poststrata with age-race groups (model assumption).
4. Determine the parameter so that resulting estimates (which correct for correlation bias) for male poststrata aggregate to the control totals from step 2.
5. The measures of difference between these estimates and the usual DSEs estimate correlation bias for adult male poststrata.

Estimation of correlation bias—the "two-group model"

1. Assume $E(\text{DSE}_{Fk}) = N_{Fk}$
2. Control total for males: $\widehat{N}_M^{DA} = (\text{DA sex ratio}) \times \sum_k \text{DSE}_{Fk}$
3. Two-group model assumption:

$$E(\text{DSE}_{Mk}) = \eta N_{Mk}$$

   where η is constant over poststrata k (within the age-race group).
4. Estimate of model parameter η:

$$\widehat{\eta} = \frac{\text{A.C.E. sex ratio}}{\text{DA sex ratio}}$$

5. Relative correlation bias:

$$\frac{E(\text{DSE}_{Mk}) - N_{Mk}}{N_{Mk}} = \eta - 1$$

   Note this is constant over poststrata k (within the age-race group).

Results (Table 3) and conclusions for 2000 A.C.E.:

- Evidence of significant (negative) correlation bias in DSEs for adult Black males, similar to 1990.
- Estimates of correlation bias for Nonblack Males 30-49 and 50+ are negative and small in magnitude, also similar to 1990.
- DA estimates for Nonblacks 18-29 are too unreliable, both the totals *and* the sex ratios, for use in estimating correlation bias. I thus conclude we cannot estimate correlation bias for Nonblack males 18-29. I recommend we proceed with the assumption of no correlation bias for Nonblack males 18-29 on the grounds that
  - estimates of correlation bias for older Nonblack males are small in 2000, and
  - the estimate of correlation bias for Nonblack males 18-29 in 1990 was small.

Advantages to the two-group model (relative to other possible models)

- Simple, results easy to interpret
- Low variance
- Corresponds to assumption of constant relative correlation bias
- Not affected by negative cells
  - Blacks: 6 of 52 poststrata (12%, compared to 22% in 1990)
  - Nonblacks: 97 of 364 poststrata (27%, compared to 35% in 1990)
- Less subject to Wachter-Freedman criticism (discussed next)

Will later develop correlation bias estimates from other models for comparison.

What is the Wachter-Freedman criticism?

- Wachter and Freedman (1999) pointed out that when estimating correlation bias one should allow for other biases in the data (and we didn't do this in 1990).
- Spencer (2000) responded that if the other biases are the same for males and females. and correlation bias (for males) is estimated by combining DSEs with DA sex ratios, then correlation bias estimates at the national level are unaffected by the other biases.
- Spencer's argument applies only at the national level, and not directly for individual male poststrata, for most combining models. However, since all the combining models used produce estimates that agree with the same male control totals at the national level, these models produce alternative allocations of the national level correlation bias that should still be useful.
- Fortunately, other biases that are the same for males and females do not affect relative correlation bias estimates for male poststrata from the two-group model.
- Lacking evidence against the assumption that other biases in DSEs aggregated to the national level are the same for males and females, we shall make this assumption and estimate correlation bias for males from the two-group model ignoring other biases. (Note: We can make adjustments for ratio bias by poststrata. This had negligible impact on the results.)

References

[1] Spencer, Bruce D. (2000), "Final Report on Correlation Bias Methodology," report prepared under contract number 50-YABC-7-66020 for the U.S. Bureau of the Census, Abt Associates, Inc. and Spencer Statistics, Inc., April 20, 2000.

[2] Wachter, Kenneth W. and Freedman, David A. (1999) "The Fifth Cell: Correlation Bias in U.S. Census Adjustment," Technical Report Number 570. Department of Statistics. University of California, Berkeley.

Table 1.a 2000 A.C.E. Percent Coverage Differences from DA Totals[1]
100×(DA - ACE)/DA

Age	Black[2] Male	Black[2] Female	Nonblack Male	Nonblack Female
0-17[3]	—	—	−2.5	—
18-29	—	—	−8.8	−6.7
30-49	—	—	−2.5	−2.6
50+	—	—	0.0	−.9

Table 1.b 1990 PES 357 Percent Coverage Differences from DA Totals[1]
100×(DA - PES)/DA

Age	Black[2] Male	Black[2] Female	Nonblack Male	Nonblack Female
0-17[3]	1.4	—	−.7	—
18-29	7.8	−.2	−1.8	−2.1
30-49	7.8	.1	.8	−.9
50+	8.8	.6	2.4	1.2

Notes:

1. The census count of the group quarters population, which is not part of the A.C.E. universe, is subtracted from the 2000 DA totals. Race assignment of the group quarters population for 2000 uses Model 2 (anyone who checked Black is assigned to the Black group.) Results for race assignement using Model 1 (only those who checked only Black are assigned to the Black group) differ some from those shown here, particularly for Blacks. For 1990 a smaller population was subtracted out as not part of the PES universe.

2. Results for Blacks for 2000 are currently omitted due to significant comparability problems arising from Black Hispanics, whom DA assigns to Black and A.C.E. to Nonblack. These results will be added to the tables when adjustments to make the DA and A.C.E. totals comparable are available. This comparability problem also affects the Nonblack results, but not very much, so the results shown can be taken as reasonable indicators of the general magnitudes of the DA and 2000 A.C.E. differences for Nonblacks.

3. Results shown under "Male" for 0-17 are actually for the total of males and females.

Table 2.a Sex Ratios from DA[1] and 2000 A.C.E.

Age	Black[2] A.C.E.	Black[2] DA	Nonblack A.C.E.	Nonblack DA
18-29	.84	.90	**1.05**	**1.03**
30-49	.82	.89	1.00	1.00
50+	.72	.76	.85	.86

Table 2.b Sex Ratios from DA and 1990 PES

Age	Black PES	Black DA	Nonblack PES	Nonblack DA
18-29	.83	.90	1.02	1.02
30-49	.84	.91	.99	1.01
50+	.72	.78	.81	.82

Notes:

1. For 2000, DA estimates revised as of 2/7/01 are used. The census count of the group quarters population, which is not part of the A.C.E. universe, is subtracted from the 2000 DA totals before computing the sex ratios. For 1990 a smaller population was subtracted out as not part of the PES universe. Race assignment of the group quarters population for 2000 uses Model 2 (anyone who checked Black is assigned to the Black group.) Sex ratios with race assignment using Model 1 are the same as those with Model 2 to the accuracy shown in the table.

2. Comparability problems arising from Black Hispanics, whom DA assigns to Black and A.C.E. to Nonblack, are expected to have minor effects on the sex ratios for 2000. Nevertheless, the results in Table 2.a will be revised later when adjustments to make the DA and A.C.E. totals comparable are available.

Table 3.a Relative Correlation Bias Estimates (two-group model[1]) for 2000 A.C.E. (expressed as percents)

Age	Black	Nonblack
18-29	−7.4	**2.0**
30-49	−8.0	−.2
50+	−4.8	−.9

Table 3.b Relative Correlation Bias Estimates (two-group model[1]) for 1990 PES (expressed as percents)

Age	Black	Nonblack
18-29	−8.0	−.3
30-49	−7.7	−1.6
50+	−8.2	−1.2

Notes:

1. For the two-group model the estimate of relative correlation bias, expressed as a percentage, is
$$100\left(\frac{\text{A.C.E. sex ratio}}{\text{DA sex ratio}} - 1\right).$$

2. Comparability problems arising from Black Hispanics, whom DA assigns to Black and A.C.E. to Nonblack, are expected to be minor for these results.

ESCAP MEETING NO. 36 - 02/12/01

MINUTES

Minutes of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 36 February 12, 2001

Prepared by: Nick Birnbaum

The thirty-sixth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 12, 2001 at 10:30.

The agenda for the meeting was to discuss the estimates of correlation bias in the Census 2000 A.C.E. estimates.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Carol Van Horn

Deputy Director/Acting Director:
William Barron

Other Attendees:

Marvin Raines
Tommy Wright
Donna Kostanich
Raj Singh
Gregg Robinson
William Bell
Deborah Fenstermaker
Alfredo Navarro
Mary Mulry
Michael Starsinic
Roxie Jones
Kathleen Styles
Nick Birnbaum
Sarah Brady
Annette Quinlan
Maria Urrutia

I. Estimates of Correlation Bias in the Census 2000 A.C.E. Estimates

William Bell began his presentation by providing the Committee with some background information on correlation bias – what causes correlation bias and how it is estimated. The handouts are attached. Correlation bias occurs when census and A.C.E. inclusion probabilities vary over persons within post-strata, which is called heterogeneity. This bias could also occur when there is causal dependence – the act of being included in the census makes someone more likely or less likely to be included in the A.C.E. For instance, if those more likely to be included in the census are also more likely to be included in the A.C.E., then the correlation bias would be negative – that is, the DSEs underestimate the true population.

The methodology being used in 2000 to estimate correlation bias in the A.C.E. was briefly contrasted with the methodology used in 1990. However, for both 1990 and 2000 the methods assume no correlation bias for adult females. The DSEs for females are aggregated to the national level within age-race groups. The female totals are multiplied by the DA sex ratios (M/F) to get control totals for adult males. Estimates for male post-strata are computed so that they aggregate to the control totals. The differences between these estimates and the DSEs provide estimates of correlation bias for the adult male post-strata. Sex ratios are used because they are expected to be more robust than DA estimates of levels.

The estimates of correlation bias were presented for 2000, compared to 1990 data, and discussed. For adult Black males, the results were consistent with 1990 findings; that is, they showed significant negative correlation bias. For nonblack males in the 18-29 age group, the correlation bias was high, not consistent with the 1990 data, but rather, indicative of a larger than expected overcount for this age-sex-race group. This result caused the Committee to question the estimate of correlation bias for this particular group. The Committee agreed that further discussion would be required to explain this anomalous result.

Finally, William Bell briefly discussed the Wachter/Freedman concerns regarding estimating correlation bias. After brief discussion, members of the Committee concurred that these concerns had been adequately addressed in the technical literature, and noted, in fact, that the two-group model was less susceptible (vis-a-vis other models) to these concerns.

II. Next Meeting

The agenda for the next meeting, to be held on February 13, 2001, is to examine the A.C.E. Dual System Estimate variances by geographic area and the total error model results.